Language Error in Aviation Maintenance

Final Report

C. G. Drury, J. Ma and C. V. Marin

August 2005

Prepared for:

Federal Aviation Administration
William J. Hughes Technical Center
Atlantic City International Airport, NJ 08405
Contract #2002-G-025
Dr. William Krebs

TABLE OF CONTENTS

List of Figures

List of Tables

List of Acronyms

AECMA………………………Aircraft European Contractors Manufacturers Association
AIAA……………………………...American Institute of Aeronautics and Astronautics
AMT…………………………………………………Aviation Maintenance Technician
ANCOVA……………………………………………………Analyses of Covariance
ANOVA…………………………………………………….....Analysis of Variance
ASRS………………………………………….. NASA Aviation Safety Reporting System
CACC……………………………………………..Civil Aviation Administration of China
CAUC……………………………………………......Civil Aviation University of China
FAA………………………………………………….. Federal Aviation Administration
FAR……………………………………………………….Federal Aviation Regulation
GRE……………………………………………………...Graduate Record Examination
GLM………………………………………………………….General Linear Models
JAA………………………………………….………….Joint Aviation Authorities
JAR…………………………………………………………..Joint Aviation Repair
MRO…………………………………………….……Maintenance, Repair & Overhaul
NASA…………………………………National Aeronautics and Space Administration
NRR……………………………………….....…………………Non-routine Repair
NTSB……………………………………………National Transportation Safety Board
OEM…………………………………………………Original Equipment Manufacture
PA…………………………………………………………………..Public Address
QA………………………………………….....…………………Quality Assurance
SHELL…………………………..Software, Hardware, Environment, Liveware, Liveware
TOEFL……………………………………………Test of English as a Foreign Language

EXECUTIVE SUMMARY

In the past decade there has been a great increase in contract maintenance among major airlines, to a current level of about 50%. The fastest growing segment of the worldwide Maintenance Repair and Overhaul (MRO) market is outside the USA. Although English is the language of aviation, it is certainly not the native language of most of the world. Thus, language errors may well arise for maintenance of the US civil fleet due to non-native English speakers interacting with maintenance materials in English. This study assists the Federal Aviation Administration (FAA) in answering a Secretary of Transportation recommendation that:

> *"The FAA should establish a method for determining whether language barriers result in maintenance deficiencies."*

The contribution of this study has been to collect quantitative evidence to test whether language errors pose a problem for maintenance and further to provide quantitative evidence for how any potential problem can be managed. A total of 941 aviation maintenance workers on four continents were tested to measure the incidence of language error, to examine the factors leading to such error and its detection, and to measure the effectiveness of chosen interventions.

From analysis of an Original Equipment Manufacture (OEM) database on language use, we were able to find the prevalence of English and native language usage from 113 airlines around the world. English verbal abilities were highest in North America, followed by Europe and lowest in Asia and the rest of the world. Translation of maintenance manuals was rare while translation of task cards and engineering orders was more common in Asia. Most meetings and training were conducted in the native language. Even among airlines with low reported levels of English ability, translation of documents was uncommon.

Our own observations and focus groups in USA and UK helped develop a pattern of language error scenarios, and a set of factors that may influence the frequency of these scenarios. The frequency of occurrence of these scenarios, and factors affecting their incidence and mitigation, was measured in the study of 941 maintenance personnel, largely Aviation Maintenance Technicians (AMTs). A comprehension test methodology quantified the effectiveness of language error interventions, for example by providing a bilingual coach, providing an English – native language glossary or using a full or partial translation of a document.

From the OEM survey and demographic data on language use, the choice of sites was narrowed to those using a form of Chinese or Spanish. With English, these two languages are the most commonly used on earth. We chose as regions Asia, Latin America and Europe (Spain), with a control sample from the USA. While we were measuring scenario frequency and intervention effectiveness, we also collected data on English vocabulary of participants, which gives a direct estimate of reading level on a scale equivalent to US grades in school. Finally, focus groups were held at each site to discuss how that MRO coped with the potential for language error.

Three scenarios related to AMT abilities to understand written and verbal communication in English were the most frequent, being seen by participants about 4 to 10 times per year. Most

language errors were detected early in the communications process. The reading grade level of participants at USA MROs was about 14 as found in earlier studies. For MROs on other continents, the reading grade level was about 5, with higher levels where there was a history of bilingualism. On all continents, task card comprehension performance improved with reading grade level. In that test, accuracy performance was generally good, and was better in areas that were bilingual. None of the interventions except translation proved effective. Partial translation, leaving technical terms in English, proved as effective as full translation. The use of good practices in documentation design was seen as a contributing factor to language error mitigation.

A set of practical interventions emerged from the scenario frequency estimates, the comprehension test and the focus groups. These are given in Chapter 9 as findings and recommendations. Design of work documentation is the primary way to reduce written language errors. Good design practice helps reduce errors and translation into the native language, if performed carefully, was found to increase document comprehension. Individual ability of Aviation Maintenance Technicians (AMTs), inspectors, managers and engineers in written and verbal English communication was important, and can be improved by training and controlled practice. The organizational environment should recognize the deleterious effects of time pressure on errors, and also recognize the symptoms of imperfect communication when it occurs. The organization also needs to plan work assignments to allow AMTs to become more familiar with particular tasks, and provide planned English practice for all personnel. Time pressure on tasks needs to be managed if language errors are to be reduced.

Chapter 1. INTRODUCTION: PROBLEM, MODELS AND ARCHIVAL DATA

As the Federal Aviation Administration's (FAA) Office of the Inspector General reported in June 2005, the country's air carrier industry is in an era of transition (OIG, 2005). Part of the cause of the transition is record financial losses for FAR Part 121 carriers, and part of the solution is seen as contract maintenance, popularly known as outsourcing. The volume of contract maintenance has been increasing each year, with the percentages for both large Network carriers and Low Cost operators now exceeding 50%. Indeed, an Analysis of Variance of the percentage data (quoted on page 8 of OIG, 2005) for 2002-2004 showed a significant year effect ($F(2,22) = 6.04$, $p = 0{,}008$), with the mean data shown in Figure 1-1. For comparison, the percentage outsourcing in 1996 was 37% (OIG, 2003).

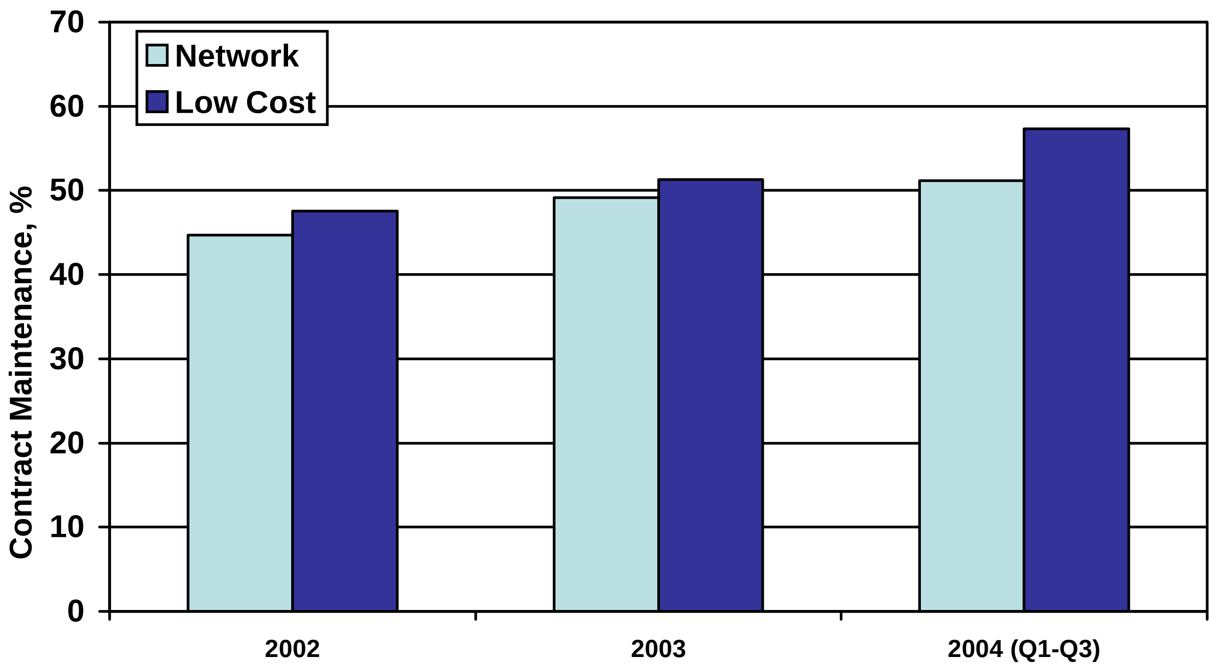


Figure 1-1. Mean percentage contract maintenance for network and low cost airlines from 2002 to 2004, from raw data in OIG (2005)

Contract maintenance is a preferred corporate strategy for reducing nonessential costs and focusing an organization on its core business (Cant and Jeynes, 1998). In aviation maintenance, Contract maintenance has been advocated and widely used, as it avoids tying up capital in maintenance facilities, and can reduce costs by opening the airline's maintenance operation to outside competition. One potential impact of such Contract maintenance is that there are more interfaces within the system, each of which represents an opportunity for error. The "system" without Contract maintenance includes the aircraft itself, the airline and the regulatory agency (e.g. the FAA). However, with Contract maintenance, a fourth organization is added to the system: the Maintenance/ Repair Organization (MRO). Drury, Wenner and Kritkausky (2000) provided models of these interactions and examined potential and actual error sources from using MROs. Data collection at a number of domestic and foreign MROs did indeed show a potential for increased errors, but little evidence of errors in practice.

Sparaco (2002) sees the formation of global MRO networks involving US and foreign airlines, as well as repair stations. In addition to offshore MROs, there are many within the USA where non-native English speakers form part of the labor pool. The difficulty of moving between languages creates an additional potential for error. The language of aviation is primarily English, both in operations and in maintenance. Aviation Maintenance Technicians (AMTs) must pass their examinations in English, and maintenance documentation in use at the Federal Aviation Administration (FAA) approved facilities is in English. This poses a second-language or translation burden for Non-Native English Speakers (NNESs) that can potentially increase their workload, their performance time or their error rate, or even all three measures.

In a 2001 report to the Secretary of Transportation by the Aircraft Repair and Maintenance Advisory Committee, many of these issues were raised in considering changes to the domestic and foreign FAR Part 145. They recommended that:

> *"The FAA should establish a method for determining whether language barriers result in maintenance deficiencies."*

This project is a direct response to these concerns that NNES, in repair stations in the USA and abroad, may be prone to an increased error rate that could potentially affect airworthiness.

1.1 Models of Communication

Communication is defined as "a dynamic and irreversible process by which we engage and interpret messages within a given situation or context, and it reveals the dynamic nature of relationships and organizations" (Rifkind, 1996). Communication can be formal or informal. Davidmann (1998) made a distinction between formal and informal communication, where formal communication implies that a record is kept of what has been said or written, so that it can be attributed to its originator. On the whole, written communications are formal. Most on-the-job communication is informal, unwritten, and sometimes even unspoken. An important distinction made in communication theory is the temporal aspect: communication is either synchronous or asynchronous. In aviation maintenance, synchronous communication is typically verbal, e.g. conversations or PA announcements, while asynchronous communication is typically written, e.g. work documentation or placards. In the context of aviation maintenance and inspection, communication has been the most frequent aspect studied since the human factors movement began there in the early 1990's (Taylor and Patankar, 2000).

The fundamental function of communication is to deliver a message from one human being to another. In almost every aspect of aviation work, communication also fulfills a secondary role as an enabler (or tool) that makes it possible to accomplish a piece of work (Kanki and Smith, 2001). Based on examination of accident investigations and incident reports, Orasanu, Davision and Fischer (1997) summarized how ineffective communication can compromise aviation safety in three basic ways:

1. Wrong information may be used.
2. Situation awareness may be lost.
3. Participants may fail to build a shared model of the present situation at a team level.

Communication models in the form of generally simple diagrams are important in helping people to understand the concept and process (Wideman, 2002). Kanki and Smith (2001) state that human communication always takes place within a set of contexts, such as a social context, a physical context and/or an operational context. Compared to some other work settings, the aviation operational context is relatively structured by standard operating procedures that organize task performance. Figure 1-2 presents a communication model we synthesized from our literature review.

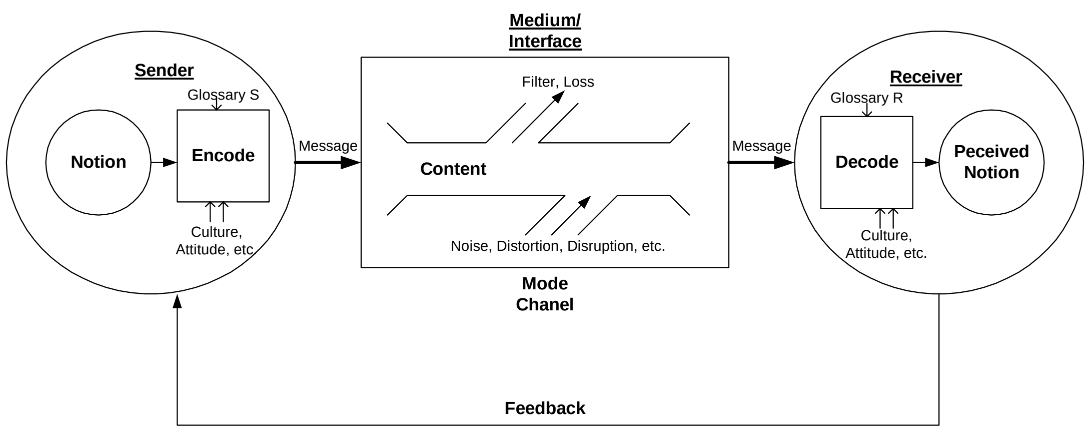


Figure 1-2. The Communication Model synthesized from literature review (Wideman, 2002; McAuley, 1979; Johnson, 1972,etc.)

Based on basic communication theories, a communication process is composed of *the sender/receiver* (e.g. people, manuals, computers, etc.), *the message* (e.g. information, emotions, questions, etc.), the *medium* (e.g. speech, text, sensory, etc.), *filters/barriers*, *feedback*, etc. (Kanki and Smith, 2001; Griffith, 1999).

Fegyveresi (1997) summarized many variables that influence communication, such as workload, fatigue, personality traits, gender bias, standard phraseology, experience level, vocal cues, etc. Language and cultural diversity can intensify differences and confusions in communication, but a language barrier does not necessarily result in unsafe cockpit operations (Merritt and Ratwatte, 1997). In order to eliminate or at least minimize potential ambiguities and other variances, people establish rules regarding which words, phrases, or other elements will be used for communication, their meaning, and the way they will be connected with one another. The aggregation of these rules is known as a “protocol.” There are four types of protocol related to flight and aircraft safety (Rifkind, 1996): verbal, written, graphical, and gestural protocols. According to Rifkind (1996), the only verbal protocol that has been established throughout aviation, including maintenance, is the use of English as the standard language. This was done when the International Civil Aviation Organization (ICAO) was established in 1944.

1.1.1 Current Data Sources 1: ASRS AND AIDS

Before field data is collected on language-related maintenance and inspection errors, existing databases need to be searched for relevant reports of such errors. The most useful of these were the NASA/FAA Aviation Safety Reporting System (ASRS) and the Accident/Incident Data System (AIDS). Our main interest was in maintenance and inspection errors, but few were reported in the databases studied. Hence, our objective changed to include all language-related errors, whether by flight crew, ATC, cabin crew or ground crew. This decision was in line with our literature search, which we broadened to include all communication errors. With a large enough set of aviation-related language errors, we can form more general models, of which maintenance and inspection errors will be a specific instance.

Based on a preliminary reading of about 60 incident reports, a taxonomy was developed of error manifestations, causal factors and recovery mechanisms. Some entries in this taxonomy reflect the earlier analysis by Orasanu, Davision and Fischer (1997), although we have tried to separate contributing factors from recovery mechanisms. This preliminary reading also found likely key words for searches. Two keyword searches were made of the ASRS and AIDS databases. The first was on “English” and the second on “Language.” We classified 684 incidents by *error type*, *contributing factor*, and *recovery mechanism*. Details are not presented here due to space limitations.

The main division of error types was between synchronous and asynchronous communication. Within these, a relatively fine classification was made by the roles of the two communicators, e.g. flight crew with ground crew. This classification was eventually collapsed into four categories. Note that “language” was used to refer to two items. Language could mean the actual language used (e.g. French, Spanish, Chinese, English) or the choice of words/phrases (e.g. listener expected one term but communicator used what was incorrectly thought to be a synonym). Some of the communication channels themselves were poor, classified here as low signal/noise ratio. In many cases, the report mentioned that at least one of the communicators was inexperienced, for example an American crew’s first flight for some years into a Mexican airport.

The analysis of the ASRS and AIDS databases used a cross-tabulation technique developed by Wenner and Drury (2000) to show significant and often interesting conclusions in Figure 1-3 and Figure 1-4. When the error locus was classified by the roles of the communicators, differences in contributing factors and recovery mechanisms were seen. Our four categories of causal factors gave roughly equal counts in the databases, showing that the use of other than a native language was an important causal factor in these errors. This contributing factor appeared to be distributed across error loci, except for asynchronous communication, where it was under-represented. In fact, for asynchronous communication as a whole, native language and low signal/noise ratio were under-represented factors, while unclear terminology was over-represented. For recovery, asynchronous communication had the least opportunity for recovery mechanisms. In particular, the repetition useful in synchronous communications was not usually fruitful.

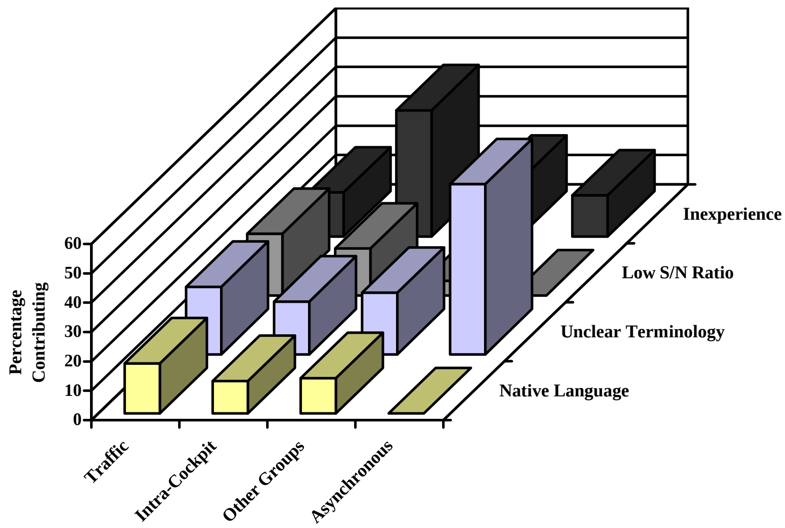


Figure 1-3. Pattern of contributing factors across error loci

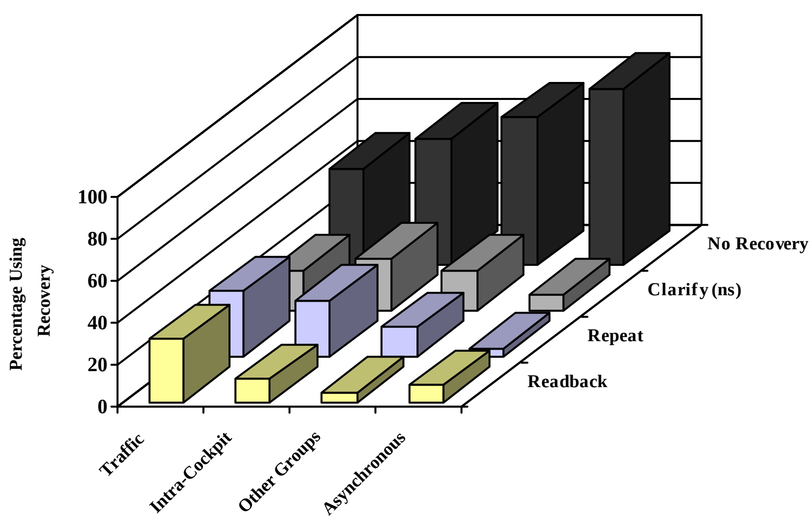


Figure 1-4. Pattern of recovery attempts across error loci

The characteristics of maintenance communications errors found here (asynchronous, terminology-related, few recovery mechanisms) helped to set the stage for our direct measurement of these errors from maintenance participant interviews and questionnaires.

1.1.2 Current Data Sources 2: OEM Survey of Airlines

From September 2002 to January 2003, an international corporation surveyed a large number of airlines throughout the world concerning their use of English and other languages in flight operations and maintenance operations. The database used was based on a large sample (n = 113) of airlines, approximately evenly divided between North America, Europe, Asia and the rest of the world. As we wished to perform statistical analyses testing for differences between regions, we needed to have enough responses in each Region, based on different challenges in language they are a priori likely to face. Table 1-1 shows the numbers in each region.

Region	Number Responding
Europe	35
North America	16
Asia	30
Other	32

Table 1-1. Regional airline sample numbers

The questions analyzed in this paper were all completed by Maintenance personnel. The first analysis was for the question on "Has your company identified any of the following issues with using the checklists in the manufacturer's handbook? Check all that apply." This question was the only one that addressed the potential errors in checklist use, so that even though it specifically referenced Flight Operations, it was deemed relevant to any study of patterns of error in communication tasks. Each response type (e.g. Difficulty finding a checklist) was tested with a Chi-square test of equality of proportions across the four Regions. Table 1-2 summarizes the result.

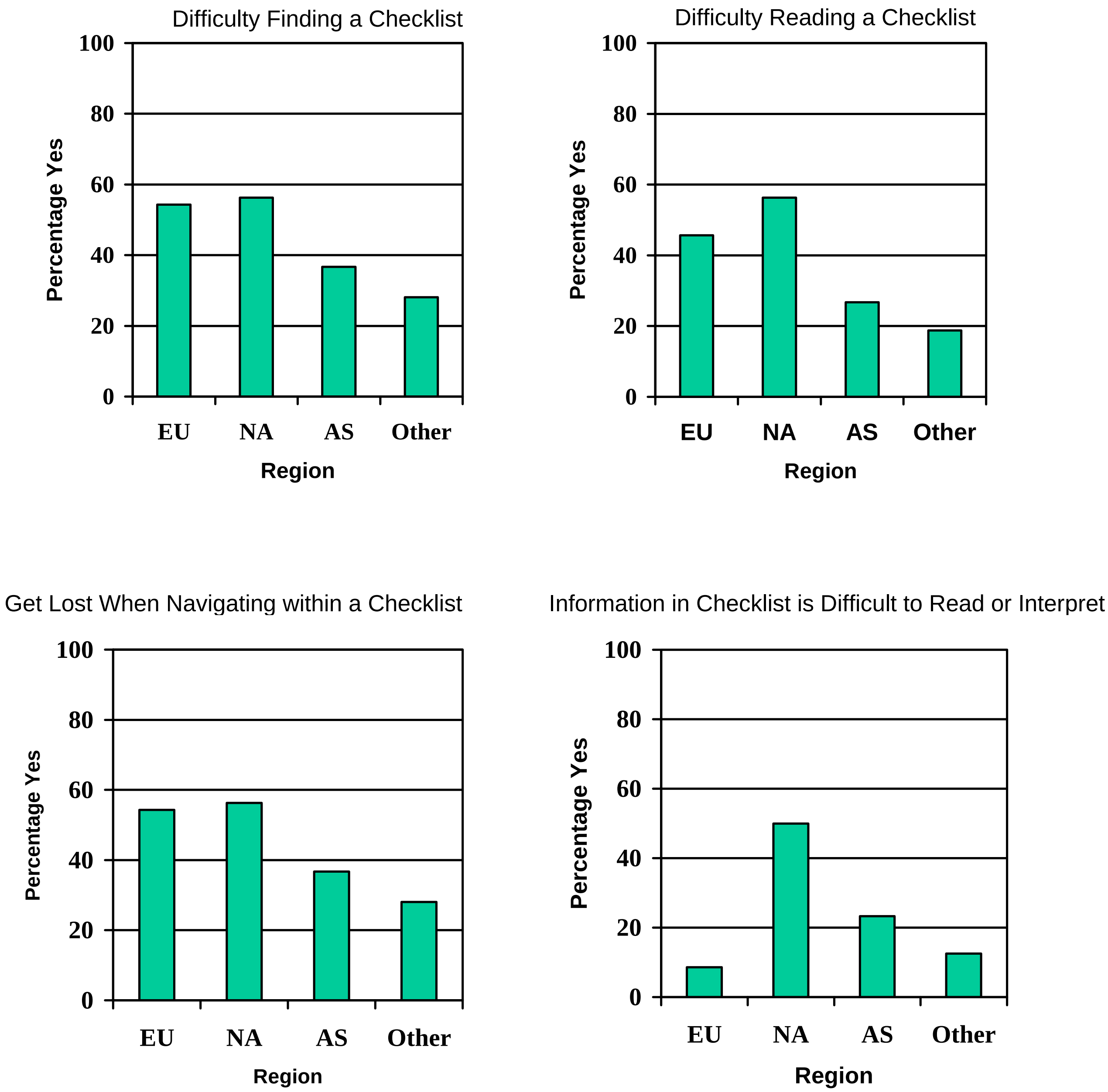
Note first that there was a considerable difference between the sixteen error patterns, with some having over twice the frequency of others. The higher frequency issues are to do with physically locating the correct checklist and performing it despite interruptions. In contrast, the lower frequency items have to do with the design, formatting and wording of the checklist itself. Clearly, the airlines in this sample reported operational rather than design difficulties with checklists.

Handbook Issue	Overall % Yes	Chi-Square	Significance
Difficulty finding a checklist	38	10.7	0.014
Difficulty reading a checklist	28	13.7	0.003
Difficulty understanding the checklist	33	4.9	Ns
Don't understand why a checklist or checklist step is needed	30	4.7	Ns
Difficulty conducting a checklist	17	2.7	Ns
Difficulty following a condition statement	32	6.1	Ns
Get lost when navigating within a checklist	46	8.2	0.042
Forget to complete a checklist item after an interruption	35	1.6	Ns
Skip a step	38	4.7	Ns
Forget to complete a checklist	23	3.0	Ns
Complete the wrong checklist	45	1.0	Ns
Difficulty confirming the checklist is the right checklist	38	3.3	Ns
Performing the checklist properly relies too much on pilot's system knowledge	19	3.4	Ns
The actions each pilot should perform or not perform are unclear	26	1.4	Ns
Not enough information to support crew in conduct of checklist	18	2.2	Ns
Information in checklist is difficult to read or interpret	21	11.8	0.008

Table 1-2. Difficulties in handbook use reported by region

For the four issues where there was a significant difference between regions, Figure 1-5 shows these differences. For the first three issues (finding, reading, and navigating) Europe and North America reported the most instances, with Asia and Other progressively less. There should be no regional differences in at least finding a checklist, so that differential willingness to report may be an issue here. Similarly, the final graph of Figure 1-5 on difficulty of reading and interpreting information on the checklist, has North America showing almost twice as many responses as any other region. Again, one would expect *less* difficulty understanding checklist information in a predominantly English-speaking population, so perhaps the North American users are less tolerant of sub-optimal conditions than those regions where an implicit translation from English to a native language is often required. This may also explain to some extent that checklist design problems are less likely to be reported than operational problems.

Figure 1-5. Significant Differences between Regions of Checklist Usage



Our next analysis was of the reported English language ability of mechanics. One question asked "Estimate the percentage of your mechanics who are described by each of the following levels of English speaking ability". Four levels of ability were presented:

- Can speak or understand very little English
- Can speak or understand a few English words or phrases
- Can speak and understand English for simple conversations
- Can speak and understand English in long, complex conversations

Similarly, the next question asked "Estimate the percentage of your mechanics who are described by each of the following levels of English reading ability". Three levels of ability were presented:

- Can read very little English
- Can read English for maintenance documents
- Can read any English document

The data from each level was analyzed separately using one-way ANOVAs between the four Regions. For Speaking ability all levels showed significant differences between regions as given in Table 1-3.

English Speaking Level	F(3,84)	p
Can speak or understand very little English	4.53	0.005
Can speak or understand a few English words or phrases	10.12	<0.001
Can speak and understand English for simple conversations	7.06	<0.001
Can speak and understand English in long, complex conversations	16.73	<0.001

Table 1-3. Regional differences between English speaking levels

Figure 1-6 shows graphically how the speaking ability varied between Regions. Note the contrast between Europe and North America, where most of the mechanics speak English, and Asia and Other, where there is less reported speaking ability.

Exactly similar analyses were performed for English reading ability where again all levels showed significant differences between regions shown in Table 1-4.

English Reading Level	F(2, 82)	p
Can read very little English	7.08	<0.001
Can read English for maintenance documents	7.42	<0.001
Can read any English document	10.81	<0.001

Table 1-4. Regional differences between English reading levels

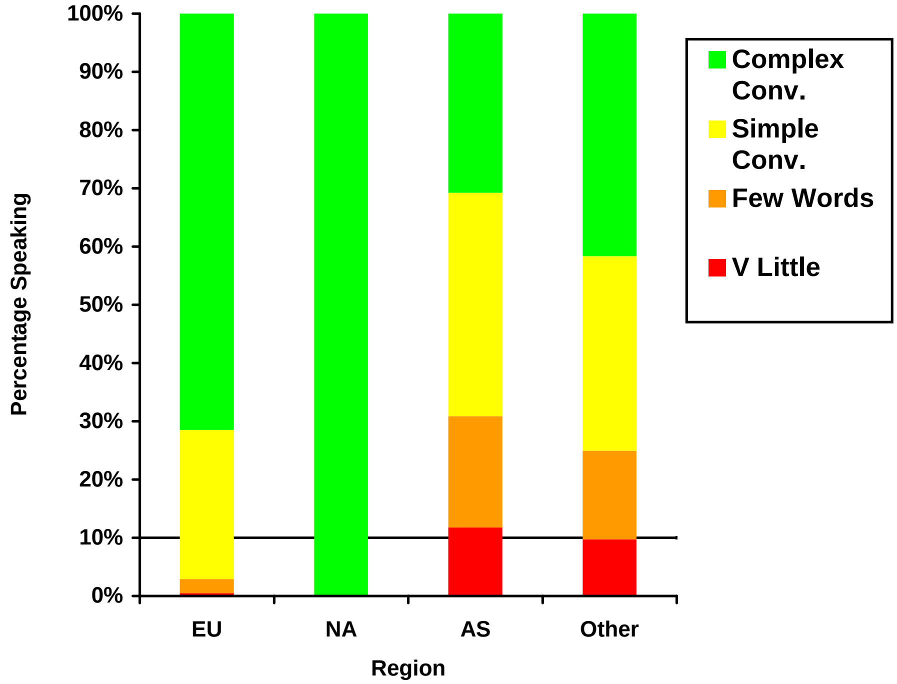


Figure 1-6. English speaking ability reported by Region

These differences are shown in Figure 1-7 with common scales to provide comparisons. Note again that there is a difference in English abilities between the two groups (Europe, North America and Asia, Other).

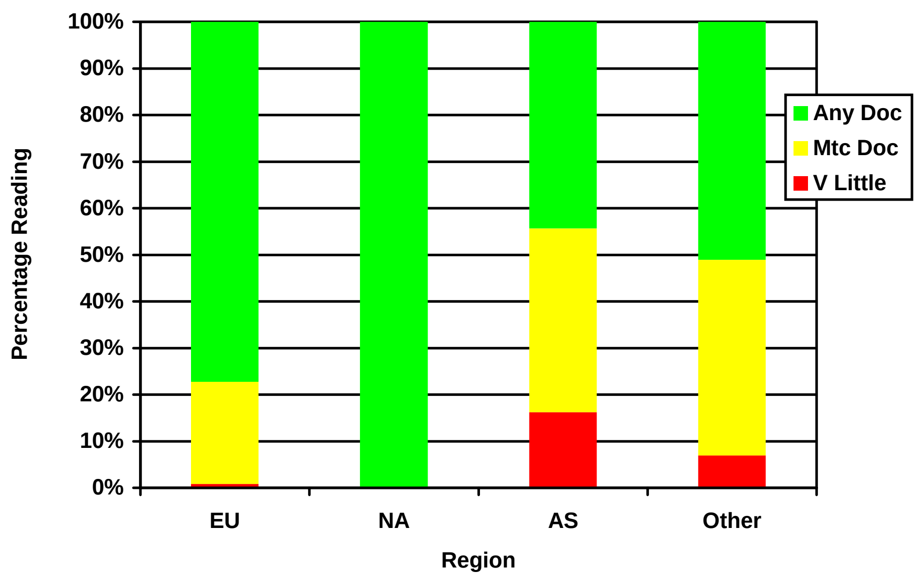


Figure 1-7. English reading ability reported by Region

The next analysis was for the techniques used in each airline to deal with consequences of any differences between the English language of documents and the native language of mechanics.

Three questions asked about translation of four types of document: Maintenance Manual, Maintenance Task Cards, and Structural Repair Manual respectively. First, data from these questions were analyzed using a Chi-square test with three responses:

- OEM document; in English and not modified
- OEM document; modified or rewritten in English by your company
- Translated to a different language

Where there were too few responses other than the first, the Chi-square was invalid and so data were recoded to give just two levels:

- OEM document; in English and not modified
- Any modification

For the Maintenance Manual, only two of the 88 airlines reported other than "OEM document; in English and not modified", with one reporting each modification. Clearly, the Maintenance Manual is typically left as supplied. For the other two documents, there was a significant difference between regions (Table 1-5) when the data were combined to two levels as above:

Document	Overall % Modified	Chi-Square	Significance
Maintenance Task Cards	13.6	26.1	< 0.001
Structural Repair Manual	4.6	12.6	Too few data

Table 1-5. Regional differences for modifications of OEM documents

In both cases the main difference was that only Asia and Other made modifications for the Maintenance Task Cards; nine airlines modified them in English while three (two in Asia, one in Other) translated them into the native language. Note that for Asia, 33% of the task cards were modified. For the Structural Repair Manual, Asia was the only Region to modify the document, with two performing each modification for a total of 13% modifying the OEM originals in some way.

The next questions all dealt with languages used in common situations. Again, the languages specified were English and a language other than English, but a mixture was also allowed if respondents checked both alternatives. The following results (Table 1-6) were obtained from Chi-square tests, with either three categories or recoded into two categories as for Question 1.

Engineering Order (2)	21.6	34.5	< 0.001
On-site maintenance training (3)	43.2	36.2	< 0.001
Meetings (3)	52.3	30.3	< 0.001
Casual talking (2)	59.1	30.1	< 0.001

Table 1-6. Regional differences for language used in different situations

Note that there were many uses of languages other than English, particularly for verbal communication. This is in contrast to the previous question where there were few translations of documents. It is more in line with Figures 6 and 7 in terms of English language abilities. The regionRegion differences are shown in Figure 1-8, where the four graphs are again combined to show similarities. As expected, North American airlines show the least use of languages other than English, with only a single airline showing a mix of English and other languages. Europe also does not use languages other than English even half the time, presumably because of the widespread use of English in the European Union countries, as well as one whole country speaking English as the primary language. However, Asia and Other regions make considerable use of languages other than English in meetings and casual talking between mechanics, with over 79% using this strategy. Asia does translate Engineering Orders most of the time, but Other regions make less than 20% use of this strategy.

We conducted a final analysis to further investigate the relationship between an airlines' actual English ability (reading and speaking) and its strategies of handling OEM English documents and oral conversation in daily maintenance practice. For actual Reading English ability, we recalculated responses to give a level of English reading and speaking ability from 0 = Very Low to 4 = High.

We expected those airlines with low level of Reading English ability would adopt some mitigating strategies in using the OEM documents (i.e. modification into AECMA simplified English, translation into their native language). However, when using the Maintenance Manual, 7 out of 8 kept the original OEM document in English without any modification or translation, while only one airline modified/rewrote it in English. When using the Structural Maintenance Manual, 6 out of 8 airlines did not make any modification or translation. Figure 1-9 demonstrates the details how these airlines deal with manufacturer's English documents.

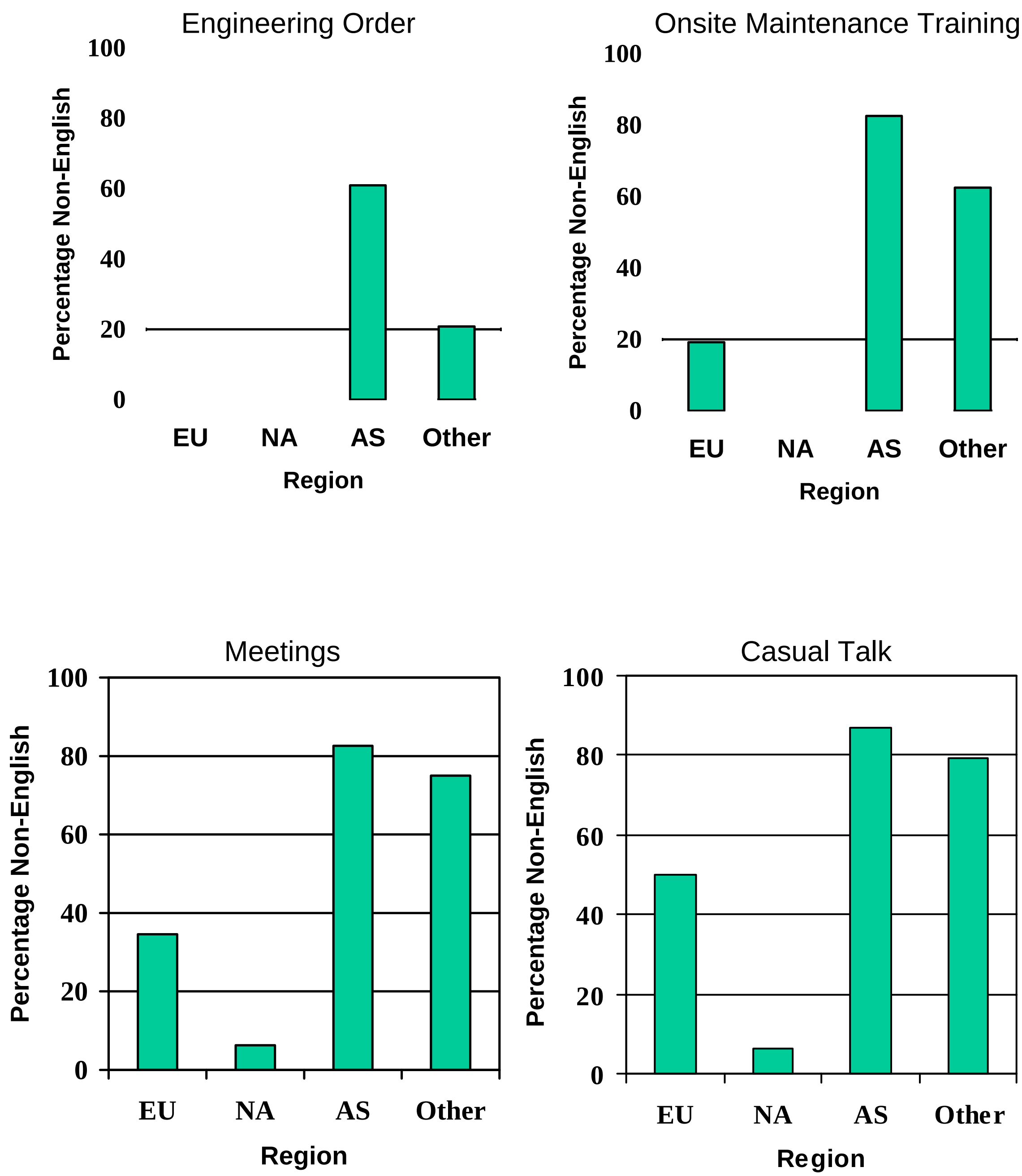


Figure 1-8. Regional difference of English usage

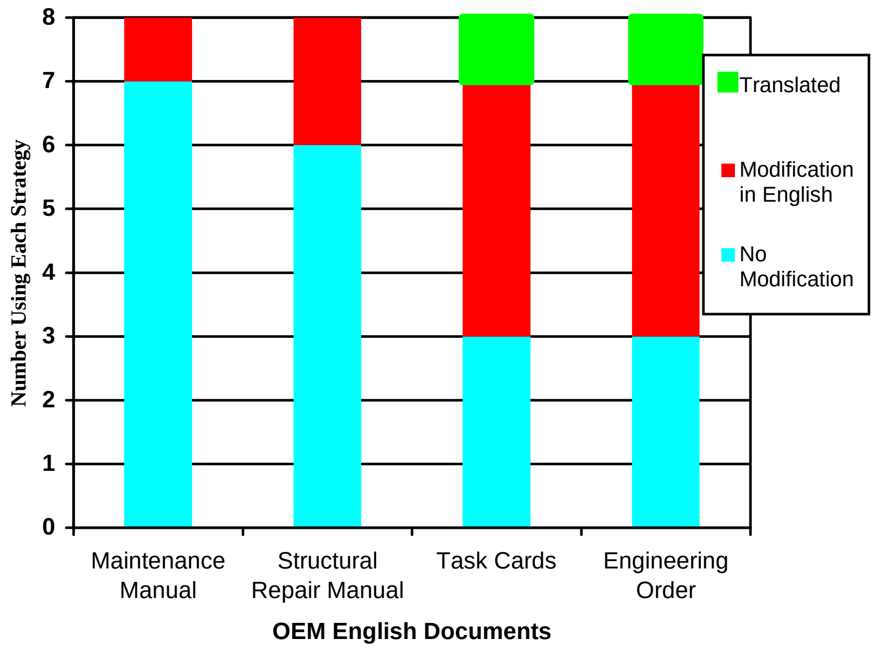


Figure 1-9. The airlines with low level of English reading ability used different strategies in handling OEM English documents

In a similar way, we analyzed the relationship between an airlines' actual ability at Speaking English and its strategies of handling oral conversation in daily maintenance practice. For actual Speaking English ability, we calculated from the answers to Q14.2.4-Q14.2.7 as following:

For those airlines with low level of Speaking English ability (categories 1-3), when conducting Onsite Maintenance Training, 100% conducted the training in a language other than English (i.e. the native language). In Meetings, 10 out of 12 airlines used another language, with the remaining two used both English and another language. Again, during Causal Talking, none of the airlines used English. Figure 1-10 demonstrates the details how they use different strategies in dealing with daily oral conversation.

Analysis of the use of English in written and spoken communications showed that English is spoken and read at a high level in North America, and to a large extent (75% or so) in Europe. In contrast, Asia and the other countries have about 50% of users able to work with written English effectively, and about 30-40% able to work with spoken English in the same way. The data from each level of English Speaking/Reading ability were analyzed separately using one-way ANOVAs among the four regions. All levels showed significant differences between regions.

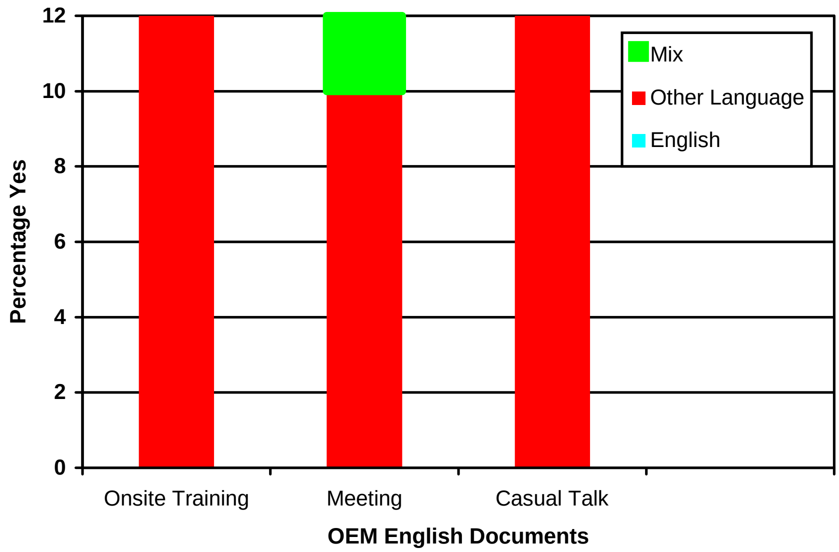


Figure 1-10. The airlines with low level of English Speaking ability used different strategies in handling daily oral conversation

The airlines cope with any potential problems through a number of means, including document translation, and conducting training and meetings in native languages. In Europe and North America, such strategies were infrequently used, presumably because most mechanics speak English, even if that is not their native language. In contrast, Asia and the rest of the world make significant use of these strategies. Translation of documents was not a common strategy, except for Asia, where 17% of airlines translated Task Cards and 60% translated Engineering Orders. Comparable figures were about 4% and 20% of airlines in other parts of the world, and almost nobody translated the Maintenance Manual. The strategy of using the native language in speaking was widely seen, with almost all Asian airlines and most airlines in other non English-speaking countries conducting meetings and maintenance training in languages other than English. However, this may represent a mismatch to documentation used in the same task that typically remained in English.

We expected that those airlines with low levels of English-reading ability would adopt some mitigating strategies in using the original documents (i.e. modification into AECMA Simplified English, translation into their native language). However, the overwhelming majority used English documents. For those airlines with a low level of English-speaking ability, almost all conducted Onsite Maintenance Training and Meetings in a language other than English (i.e. the native language). Again, during Casual Talking, none of the airlines used English.

1.3 Focus Groups on Language Errors

While the analysis of archival data in the above section could provide some insight into language errors in maintenance, such data were not collected for that purpose (c.f. Drury 1995). More direct data collection involves the use of questionnaires and interviews specifically on the theme of language errors in maintenance. However, before we can ask sensible questions, we must have valid information on the types of errors involved. We collected such data from focus groups at MROs in different countries. So far (May 2003), we have run five such focus groups, three at US-based MROs and the other two at UK-based MROs.

A focus group gathers people together to discuss the issue at hand via moderator questions and group discussions. Data are gathered through observations and conversations with participants. Focus groups are particularly appropriate for use in exploratory studies when little is known about a population or phenomenon. According to Albrecht et al. (1993), data collected in focus groups may be more ecologically valid than methods that assess individuals' opinions in a relatively asocial setting, given that language errors are social events involving the interaction of participants and the interplay and modification of ideas.

We used focus groups of people at MROs drawn from AMTs, supervisors, engineers and QA specialists. Each interview lasted about 45 minutes. Our introductory statement (after introductions, ground rules and assurance of anonymity) was:

> "We are helping the FAA to reduce errors in aviation maintenance and inspection. Our aim is to find improved ways of performing maintenance and inspection jobs. One issue has been that although English is the primary language of aviation, many people do not have English as their native language."

Then, the focus groups discussed approximately ten questions with the principal investigator as moderator. When we had transcribed the data, we compared the transcripts with our notes to look for patterns of maintenance language errors or events under four headings.

1. Error types/patterns
2. Potential error detection points in the maintenance process.
3. Factors predisposing to language errors
4. Factors potentially mitigating language errors

Function	Language Error Detection
Setup	• AMT may appear perplexed, or may agree with everything said.
Task Performance	• AMT may ask for assistance or clarification. • AMT may close access prematurely (i.e. before buyback)
Buyback	• Physical error may be detected. • AMT may not understand inspector's questions.

Table 1-7. Language errors arising in a task sequence framework

From these lists, we were able to see the functions of aircraft maintenance and inspection (see Drury, Shepherd and Johnson 1997) and where language errors could arise. Table 1-7 represents our current characterization of these situations where their errors could arise, presented within a task sequence framework. We found the following patterns of error in both verbal (synchronous) and written (asynchronous) communication:

<u>Verbal (Synchronous)</u>

1. AMT unable to communicate verbally to the level required.
2. AMT and colleagues/supervisors have poorly matched models of their own and each other's English ability.
3. Native English speakers with different regional or non-US English accents (e.g. UK, India, Caribbean) prevent adequate communications.
4. AMTs unable to understand safety announcements over the PA system.

<u>Written (Asynchronous)</u>

5. AMT unable to understand safety placard in English.
6. AMT unable to understand written English documentation.
7. Foreign documentation poorly translated into English.

These patterns form the basis for a set of seven Scenarios that were tested in the field data collection methodology defined in Chapter 2.

Table 1-8 shows the predisposing and mitigating factors identified in the focus groups. They are classified in terms of the SHELL model of human factors in aviation (Easterby, 1967).

SHELL Category	Predisposing Factors	Mitigating Factors
Software (procedures)	• Task complexity • Instruction complexity	• Document translation • Consistent terminology • Good document design
Hardware (equipment)	• Limitations of communication channel, e.g. radio, PA	• Use of aircraft as a communication device: "show me"
Environment	• Time pressure prevents AMT from querying others	
Liveware (individual)	• Inadequate written English ability • Inadequate verbal English ability • Reversion to native language under stress	• Job familiarity • Comprehension tests for AMTs • Certify AMT for specific jobs
Liveware (intercommunication)	• Unwillingness of AMT to expose their lack of English • Time pressure	• Translator available • Assign AMTs to job based on English ability • Team AMT with native English speaker

Table 1-8. Predisposing and mitigating factors identified in the focus groups

1.4 Discussion

The first phase was to find the patterns of language errors, provided there is evidence that they exist. Our analysis of communication models and the company database has shown the potential for language errors by showing that responses to language differences may not always keep pace with the need for such interventions. The ASRS database analysis showed some actual language errors, although these were mainly in the flight operations domain more likely to be reported to ASRS. Patterns in this data showed that maintenance language errors were largely asynchronous, were related to terminology and had few recovery mechanisms.

The five focus groups refined our conclusions. We now have ample evidence that language errors exist, but also that recovery mechanisms and mitigating factors are possible. The patterns found were numerous, and certainly not limited to asynchronous communication. Although documentation was an important source of difficulty, there were other patterns in verbal communication, including unexpected ones of regional accents of native English speakers. We were also able to further document the time course and propagation of errors, including error detection points and interventions. In an industry as heavily regulated as aviation maintenance, there are a number of barriers to error propagation (c.f. Reason, 1990), including the initial work assignment and inspection by a different person.

The characteristics of language errors found so far in maintenance suggests that a few overall patterns may account for most of the potential errors. In the next chapter this project, we use the patterns to design a study for collecting field data to estimate the prevalence of the patterns. It is important to have data collection in several regions of the world, for example those used in our analysis of the company database. A methodology of comprehension tests of workcards (e.g. Chervak, Drury and Oullette, 1996; Drury, Wenner and Kritkausky, 1999) is available to test the effectiveness of intervention strategies. These include use of Simplified English, full translation, use of an English-speaking coach and provision of a local language glossary. In this way, we provide a quantitative basis for recommendations to both MROs and regulatory bodies for the effective reduction of language errors.

Chapter 2. METHODOLOGY

Three aspects of interest formed the basis for our data collection efforts, designed specifically to answer FAA questions about the nature and frequency of language errors and possible interventions to reduce these errors.

First, typical demographic measures were collected for each participant. In addition we collected language-related data to characterize the population of AMTs and also to provide potential covariates for our analyses of intervention effectiveness. These were Years Studying English, and a measure of reading grade level from the Accuracy Levels Test. Second, a questionnaire was given for each scenario asking whether the respondent had encountered that scenario, how long ago, and what factors were associated with the scenario. Third, the set of interventions noted above was tested using a task card comprehension measure to find their effectiveness. Finally one or more focus groups were held at each site to better understand the way in which potential language errors were handled in their organization.

2.1 Measures

Demographic data were collected as: Age, Gender, Job Category and Years as an Aviation Maintenance Technician (AMT). The Accuracy Levels Test (Carver, 1987) is a timed 10-minute test used a total of 100 words with a forced synonym choice among three alternatives, and produced on the scale of reading grade level normed on US public schools. It has been validated against more detailed measures of reading level (Chervak, Drury, Ouellette, 1996).

Frequency measures were collected for each of the following scenarios:

> **Scenario 1:** "The Mechanic (Aircraft Maintenance Technician, AMT) or Inspector was not able to communicate verbally to the level required for adequate performance."
> **Scenario 2:** "The Mechanic (AMT) or Inspector and the person to whom they were speaking did not realize that the other had limited English ability."
> **Scenario 3:** "Native English speakers with different regional accents did not understand each others' communications."
> **Scenario 4:** "The Mechanic (AMT) or Inspector did not understand a safety announcement over the Public Address (PA) system."
> **Scenario 5:** "The Mechanic (AMT) or Inspector did not fully understand a safety placard."
> **Scenario 6:** "The Mechanic (AMT) or Inspector did not fully understand documentation in English, for example a Work Card or a Manual."
> **Scenario 7:** "The Mechanic (AMT) or Inspector did not fully understand a document translated from another language into their native language."

For each of these seven scenarios the incidence questionnaire asked first whether each had ever been encountered. This was the primary incidence measure, i.e. percentage incidence of each. To gather more detail on frequency, respondents were asked whether the scenario occurred in the past week, month, year or longer. We also asked how many months or years, but the data were not always given in a quantitative manner, so an estimate of the median time since previous occurrence was derived from the week/month/year data. Also, for each scenario, participants were asked to check the factors associated with increased likelihood of the error occurring (9 factors), with mitigating each error (10 factors) and with the discovery of each error (6 factors). The factors came from our previous analyses of databases of errors and focus groups used to derive the scenarios (Drury and Ma, 2003).

The factors for each of Error Likelihood were:

	Likelihood Factor
1	The task is complex
2	The task instructions are complex
3	The communication channel, e.g. radio or PA, interferes with good communication
4	Time pressure prevents the mechanic (AMT) or inspector from asking other people for help
5	The mechanic (AMT) or inspector has inadequate written English ability
6	The mechanic (AMT) or inspector has inadequate verbal English ability
7	The mechanic (AMT) or inspector reverts to their native language under stress
8	The mechanic (AMT) or inspector is unwilling to expose their lack of English
9	Time pressure makes the mechanic (AMT) or inspector hurry

The factors for each of Error Prevention (Mitigation Factors) were:

	Prevention Factor
1	The document is translated into the native language of the mechanic (AMT) or inspector
2	The document uses terminology consistent with other documents
3	The document follows good design practice
4	The mechanic (AMT) or inspector uses the aircraft as a communication device, for example to show the area to be inspected
5	The mechanic (AMT) or inspector is familiar with this particular job
6	The mechanic (AMT) or inspector has taken and passed a comprehension tests
7	The mechanic (AMT) or inspector was certified for that specific job
8	There is a translator available to help the mechanic (AMT) or inspector
9	Jobs are assigned to the mechanic (AMT) or inspector to job based on English ability
10	The mechanic (AMT) or inspector is teamed with a native English speaker to perform the job

The factors for each of Error Discovery were:

	Discovery Factor
1	The mechanic (AMT) or inspector appeared perplexed.
2	The mechanic (AMT) or inspector agreed with everything that was said.
3	The mechanic (AMT) or inspector asked for assistance or clarification.
4	The mechanic (AMT) or inspector closed access prematurely (i.e. before buyback)
5	The physical error resulting from the language error was detected.
6	The mechanic (AMT) or inspector did not understand inspector's questions at buy-back.

To test for how potential documentation errors can be reduced, we measured the effectiveness of document comprehension. In the study, one task card was given to each participant with a 10-item questionnaire to test comprehension. The methodology had been validated in our previous research (e.g., Chervak, et al., 1996; Drury, Wenner and Kritkausky, 1999). The comprehension score was measured by the number of correct responses, with time taken to complete the questionnaire as an additional measure. In addition, the task card was rated by the participant on the fifteen scales originally developed by Patel et al (1994).

2.2 Task Cards

We selected two task cards, one "easy" and one "difficult," from four task cards used in our previous research (Drury, Wenner and Kritkausky, 1999), because it had already been found that task difficulty affected the effectiveness of one strategy, Simplified English. As was expected, the use of Simplified English had a larger effect on more complex task cards (Chervak and Drury, 2003). The complexity of these task cards was evaluated by Boeing computational linguists and University of Washington technical communications researchers considering word count, words per sentence, percentage passive voice, and the Flesch-Kincaid reading score. The cards differed on all measures. Note that both cards were comparatively well-written, certainly compared to earlier task cards tested by Chervak et al (1996).

Both of the task cards were then prepared in the AECMA Simplified English versions, which were also critiqued by experts from Boeing, the University of Washington, and the American Institute of Aeronautics and Astronautics (AIAA) Simplified English Committee.

2.3 Experimental Design for Comprehension Test for Non-English Speaking Countries

A fully nested (between subjects) 2 × 2 × 5 design was used with factors as follows:

Task card Complexity: 2 levels - Simple
- Complex

Task card Language: 2 levels - Simplified English
- Not Simplified English

Language Interaction: 5 levels - No intervention (English)
- English with glossary
- English with coach
- Full Native Language translation
(- Partial Native Language translation)

We have already observed the use of English coaching by a more senior person, e.g., lead, foreman, engineer. Also, from the airline survey, we knew that some organizations translate documents into the native language of the employees. Finally, we have seen glossaries of English/native language words pertaining to aviation maintenance. All three are managerial practices we wished to examine through experimentation.

The Partial Native Language translation intervention was only used in Latin America and Europe as the need for this was only apparent after Asian data collection. Linguistic studies of bilingualism and sociolinguistics have illustrated a phenomenon called “Code Mixing” or” Code Switching”, i.e. the use of two languages in a conversation or in a written text (Luna and Perachio, 2005; Montes-Alcala, 2000). People are motivated to switch from one language to another to accommodate to different environments, external pressures or to align themselves with groups that they wish to be identified (Gardner-Chloros, 1991). Table 2-1 shows eight the different types of code switching that have been identified by Montes-Alcala (2000) and how these are used by people in different contexts.

Code Switching Type	Definition/Purpose/ Use
1. Quotes	To quote the original words of a person in a specific language or both.
2. Emphasis	To highlight a word, a sentence or an idea. An author can get more attention to a word by switching it to a different language.
3. Clarification or Elaboration	A word or a sentence is repeated in both languages to further develop an idea.
4. Lexical Need	A particular word that has a special meaning or connotation in one language and it does not have it (or it does not exist) in the other language, e.g. A culturally specific word or technical words like “Internet”
5. Parenthetical Comments	An author can enclose a comment in an imaginary parenthesis by writing it in a different language.
6. Idioms/ Linguistic Routines	To express an idiomatic expression or linguistic cliché.
7. Stylistic	Contributes to the rhythm of the stanzas and adds color and life to the poems
8. Free Switching	The code switching does not follow a specific pattern to make the dialogues more lively and believable.

Table 2-1. Code Switching Types (Montes-Alcala, 2000)

In the context of language errors in aviation maintenance, the only one that relay concerns us is #4: Lexical Need. Aviation English has many unique words, with new ones introduced whenever new technical systems are implemented. AMTs learn the meaning of these new words in context, on the job training and from the figures and graphs that illustrate the function of the technical term. Many times these technical words do not have a simple translation on the native language, so that it is easier to use the original term than to make an often cumbersome literal translation. This direct use of foreign terms in aviation language was common in English in the early days of aviation, where many terms were taken directly from the original French. Words such as aileron, fuselage, empennage and longeron have become so accepted into English that aviation users no longer see them as foreign imports.

When we found this phenomenon in the Asia data collection, we introduced a new intervention called partial translation to test the impact on comprehension of using original English words when the main sentences are kept in Spanish.

We originally intended to use an additional intervention using glossary plus coaching to form a 2 glossary × 2 coaching sub-design. However, in both our Chinese Engineering Graduate sample and early tests at MRO sites, it became obvious that very few participants actually used these job aids (see Section 3, Pilot Tests for more details). Thus, the glossary and coaching intervention was dropped from the study. Additionally, the Chinese and Spanish translation of each task card was only performed once, whether for Simplified English or not, so that no difference was expected between task card language for that intervention.

2.4 Choice of Participants and Sites

As noted earlier, data from the manufacturer's survey indicated that Asia, Europe and Other Countries had quite different responses. Also, recent data (Seidenman and Spanovich, 2004) suggests that Asia and Latin America are the most frequent countries for contract maintenance. Thus our choice of regions was made as Asia and Latin America, with one country from Europe and a control sample from USA.

2.4.1 Asia

The political status of the three "countries" selected, Mainland China, Hong Kong and Taiwan is complex, so they will be referred to in this report as "Areas" to avoid the impression that they are, or are not, parts of the same country. There are several reasons to collect data from MROs located in Asia, especially mainland China, Taiwan Hong Kong. First, in our analysis of the manufacturer's survey data, we found that about 30% of users in Asia had a very limited English speaking ability, another 40% were able to conduct simple conversations; about 40% of the users were able to work effectively with only written maintenance/inspection related documents, and another 15% had very little English reading ability. Compared with North America and Europe, Asia has a much smaller base of English-using mechanics. Second, the Asia-Pacific region is poised to be one of the strongest growth engines for the foreseeable future for the maintenance, repair and overhaul industry (*Overhaul & Maintenance*, 2002). U.S. and European airlines continue to ship wide-body aircraft to East Asia to take advantage of low labor costs. Almost half of the top ten Asian MROs are located in China. According to *Aviation Week & Space Technology*, "the Civil Aviation Administration of China (CAAC) is confident that despite the downturn in the global airline industry, more maintenance, repair and overhaul (MRO) joint venture companies will be set up with Chinese airlines within the next two years" (Dennis, 2002). Asia is expected to grow worldwide MRO market share from 18% in 2004 to 22% in 2014 (Seidenman and Spanovich, 2004) and is also likely to provide a wide range of managerial practices for handling language differences. By choosing China, Hong Kong and Taiwan, we were able to provide a reasonably consistent language use, and also to include an area (Hong Kong) with a tradition of English/Chinese bilingualism.

2.4.2 Latin America

Central and South America are the recipients of an increasing share of worldwide maintenance from North America (Philips, 2005). Their labor rates are generally lower than in North America (Phillips, 2004). In the OEM survey, Latin America was classed as part of "Other" regions for statistical convenience, but that whole class had similar levels of English ability to Asia. Like Asia, a single language, Spanish in this case, is predominant, making translation and administration of the data collection instruments simpler and more consistent. Within Latin America, we chose Mexico as it has a large concentration of full-service MRO sites, Colombia and Argentina as areas with much growth potential and Puerto Rico as an officially bilingual area for comparison with traditionally Spanish-only countries.

2.4.3 Europe

The second largest share of worldwide MRO market is Europe, which is expected to grow slightly from 25% in 2004 to 27% in 2014. In our OEM survey, Europe had the largest fraction of good English readers of any region, with few modifications made to task cards. Europe was expected to be the region with the best English performance, except for North America and bilingual areas such as Hong Kong and Puerto Rico. Some European-based MRO organizations, such as Lufthansa Technik, have sites in other regions as well as joint ventures with other countries (Flottau, 2005). Airlines themselves are large MRO providers, with about 60% of the current market in Europe (Phillips, 2005). Only two new comprehensive maintenance bases are expected for the Airbus A380 large transport aircraft, with one in Asia and the other in Europe (Wall, 2005). We chose Spain as we could compare the use of the same language, Spanish, to our sample of Latin American countries.

2.4.4 USA

A control group from USA was chosen to provide a baseline comparison for the other regions. We have worked with a number of airlines and MROs in the past, but this time chose an MRO that we had never tested before. The MRO had a number of sites in the USA, of which we chose two in the mid-West for convenience and for an expected low fraction of non-native English speakers.

2.5 Preparation of the Data Collection Packets

Contacts in several Chinese-speaking countries were helpful in gaining access to MROs. The translation process took place in two steps. A native Chinese research assistant (9 years as an engineering major), who is very familiar with the task cards and bilingual in English, took a lead in translating the packet. A large number of technical and language references were consulted. The principal investigator and other domain experts (e.g., native Chinese mechanical engineers in the Department of Aerospace and Mechanical Engineering at the University at Buffalo, SUNY) were consulted on the technical details (e.g., lockwire). Then both translated and original packets of data collection material were submitted to a retired professor (also fluent in English) from the Department of Avionics, Civil Aviation University of China (CAUC) for review. The translated material included the four task cards (for the full translation condition), the comprehension questions, the task card ratings, the demographic information form, the informed consent form and the questionnaire on frequency and causality of language errors.

We developed an English/Chinese glossary for each task card. We had two native English speaking engineering graduate students and two native Chinese speaking engineering graduate students read through all the task cards and circle all the words/phrases/sentences they did not comprehend, or even those about which they were slightly unsure. We built up this glossary to be as comprehensive as possible, including nouns, verbs, adjectives, abbreviations, etc.

For data collection where traditional Chinese was used (i.e., Taiwan), all forms were checked for correct current usage of traditional Chinese characters by two bilingual Chinese/English engineers with good knowledge of both human factors and aviation maintenance.

We also prepared for data collection in an Asian country with a different language, but the MROs cancelled data collection prior to our visit.

For Latin American and European countries, we established a direct contact with the FAA's liaison of each selected MRO. The selection of MROs was based on their capabilities in Airframe, Non-Destructive Test and Power plant. The translation process was done in two steps. First, a native Spanish speaking research assistant (10 years as an engineering and industrial safety majors) who is bilingual translated all the original English documents (informed consent, demographic data, task cards, glossary, comprehensive questions, partial translation, task card ratings and scenarios) to Spanish. Then, the complete package was reviewed by an expert team of a human factors engineer and two Aviation Maintenance Technicians (from Spain and Latin American countries) who currently work in a USA MRO. All suggested corrections to the translation of technical words were made until the team was satisfied.

2.6 Data Collection Process

At each MRO site, an initial meeting with management was used to explain verbally the objectives and conduct of the study, as a supplement to our earlier written communications. At this meeting, we also discussed the type of work at the site, the range of customers served and the importance of language issues and errors. Agreement was reached on the types of participants of most use to us, e.g. AMTs, engineers, QA personnel, managers. The company then scheduled multiple participants at approximately 75 minute intervals.

Groups of participants were nominally of six people, but groups with 2-10 were encountered. Each group of participants was welcomed, and the general objective of the data collection, i.e. to understand language errors and how to reduce them, was communicated. After obtaining Informed Consent and completing demographic questions, the participants all started the timed intervention evaluation (task card comprehension test) at the same time. The participants were given one of the four task cards and its associated comprehension questions. They were timed, but instructions emphasized accuracy. When this had been completed, each participant was given the rating form. The participants who finished both of these rapidly were given the seven-scenario frequency/causality questionnaire to allow slower participants to catch up. All were then given the Accuracy Levels Test, starting the 10 minute timed test at the same time. If time remained in the 75 minute session, the participants who had not completed the incidence questionnaire were given that. If there would not be time, remaining participants were asked to take that questionnaire back to their workplace and return the completed questionnaires later.

The participants were scheduled to be tested in groups with the same intervention, as far as possible. However, at times too few participants arrived so that mixed groups were sometimes tested. The participants were told that not all people in the experimental room were getting the same Task Card, or the same intervention condition. On a couple of occasions, a participant did not even attempt the task in one of the first three intervention conditions because they did not read English. In these few cases, the response was noted and the participant was given the equivalent full Chinese translation condition. We could later count this participant as scoring zero on the comprehension test if required.

The participants were assigned to the Task Card complexity and Task Card language conditions in rotation. As they were assigned to the experiment by their manager, no unwanted volunteer bias from this procedure was expected. The participants were volunteers in the experiment, but only after they had been assigned to attend by their managers.

A total of 13 focus groups, each of 6-15 engineers, quality personnel, AMTs and managers, were conducted across the sites. Discussions were wide-ranging in English and Chinese, led by one or more of the experimenters posing questions about language errors, communication problems, their causal factors and how such errors are mitigated by the organization and its people.

Chapter 3. PILOT TESTS OF METHODOLOGY

In our continuing work, we will be visiting sites worldwide to measure the frequency of these scenarios, but the current paper concentrates on the second aspect of the work, that of evaluating countermeasures.

Our analysis of worldwide survey data from a major manufacturer reported earlier found that two strategies used to reduce the potential for language errors were (a) translation into the native language, and (b) conducting face-to-face meetings in the native language. However, only about 17% of airlines in the region that most often used translation (Asia) actually translated maintenance documents into the native languages. Even among the group of 8 airlines who reported the lowest English speaking ability, only two modified the English documents in any way. Other strategies of intervention found in our site visits included having a bilingual English/native language speaker assist the mechanic with the English documentation, and/or providing a glossary of key words between the native language and English. Finally, our own earlier research into the artificial maintenance language called AECMA Simplified English (e.g., Chervak, Drury and Ouellette, 1996) had shown this to be an effective error reduction technique, particularly for non-native English speakers and for complex work documents.

Thus, we will compare four potential language error reduction interventions:

- The translation of a document into AECMA Simplified English
- The provision of a Glossary
- The provision of a bilingual coach
- The translation of a document and all related materials into a native language

Some of these methods may be combined, for example the provision of both a Glossary and a bilingual coach, or the addition of AECMA Simplified English to all conditions except for translations into the native language. Finally, for comparison, a baseline condition, no intervention, will be required. This paper describes the first two experiments conducted within this framework.

3.1 Design

As shown in Table 3-1, our study is a three-factor factorial design with the participants nested under the three factors of:

1. Task Card Complexity (Easy vs. Difficult)
2. Document Language (Simplified English vs. Non-simplified English)
3. Interventions (None, Glossary, Full Translation, Bilingual Coach, Glossary Plus Bilingual Coach)

Intervention	Easy Task Card		Difficult Task Card	
	Simplified English	Non-Simplified English	Simplified English	Non-Simplified English
	#2 / #1	#2 / #1	#2 / #1	#2 / #1
1. Control	2 / 4	2 / 3	2 / 4	2 / 4
2. Glossary	2	2	2	2
3. Tutoring	2	2	2	2
4.Glossary & Tutoring	2	2	2	2
5.Chinese Translation	2	2	2	2

NOTE: #1 represents the number of participants in Pilot Test 1, and #2 represents the number of participants in Pilot Test 2.

Table 3-1. Participant numbers by experimental conditions for Pilot Tests 1 and 2

3.2 Choice of Participants and Sites

The main task will take place at various foreign Maintenance/Repair organizations (MROs), but the two studies reported here were performed in the USA as baseline and pilot tests.

There are several reasons to collect data from MROs located in Asia, especially China. First, in our analysis of the manufacturer's survey data, we found that about 30% of users in Asia had a very limited English speaking ability, another 40% were able to conduct simple conversations; about 40% of the users were able to work effectively with only written maintenance/inspection related documents, and another 15% had very little English reading ability. Compared with North America and Europe, Asia has a much smaller base of English-using mechanics. Second, the Asia-Pacific region is poised to be one of strongest growth engines for the foreseeable future for the maintenance, repair and overhaul industry (*Overhaul & Maintenance*, 2002). U.S. and European airlines continue to ship wide-body aircraft to East Asia to take advantage of low labor costs. Almost half of the top ten Asian MROs are located in China. According to *Aviation Week & Space Technology*, "the Civil Aviation Administration of China (CAAC) is confident that despite the downturn in the global airline industry, more maintenance, repair and overhaul (MRO) joint venture companies will be set up with Chinese airlines within the next two years" (Dennis, 2002).

Participants were tested singly or in small groups. After obtaining Informed Consent and completing demographic questions, participants were given one of the four task cards and its associated comprehension questions. They were timed, but instructions emphasized accuracy. After the completion of the comprehension task, participants were given the Accuracy Level Test for the required 10 minutes. This test used a total of 100 words with a forced synonym choice among three alternatives, and produced on the scale of reading grade level. It has been validated against more detailed measures of reading level (Chervak, Drury, Ouellette, 1996).

3.3 The Preparation of the Data Collection Packet in Chinese

The translation process took place in two steps. A native Chinese research assistant (9 years as an engineering major), who is very familiar with the task cards, took a lead in translating the packet. A large number of technical and language references were consulted. The principal investigator and other domain experts (e.g., native Chinese mechanical engineers in the Department of Aerospace and Mechanical Engineering at the University at Buffalo, SUNY) were consulted on the technical details (e.g., lockwire). Then, both translated the task cards, and original packets of data collection material were submitted to a retired professor from the Department of Avionics, Civil Aviation University of China (CAUC) for a review.

We developed an English/Chinese glossary for each task card. We had two native English speaking engineering graduate students and two native Chinese speaking engineering graduate students read through all the task cards and circle all the words/phrases/sentences they did not comprehend, or even those about which they were slightly unsure. We developed this glossary to be as comprehensive as possible, including nouns, verbs, adjectives, abbreviations, etc.

With all of this material prepared, we performed two experiments before visiting the Chinese-speaking sites.

3.3 The Results of Pilot Test 1: Native English-speaking Maintenance Personnel

This test used 15 participants from three sites in the UK and the USA as a representative sample of English-speaking maintenance personnel who were unlikely to have any language errors. They were tested on the same visits where focus group data was collected, as reported in Drury and Ma (2003). All were tested under the four combinations of Simplified English/Not and Easy/Difficult Task Card to give a 2 x 2 between subjects design. There were no other interventions with these native English speakers.

First, there was a high negative correlation between accuracy and time for the comprehension test ($r = 0.692$, $p = 0.004$), and moderate correlations of both with Reading Level at $p = 0.06$. Thus, another variable was created through dividing Accuracy by Time to give a combined overall Performance score. Reading Level was tested as a covariate, but was not significant in any of three GLM ANOVAs of Accuracy, Time and Performance. In each of those ANOVAs, the only significant effect was Task Card, which was significant at $p = 0.044$, 0.012 and 0.017, respectively. As shown in Table 3-2, the difficult task card had worse performance on all variables than did the easy task card.

	Accuracy, %	Time, s	Performance, %/s
Easy Task Card	74	754	0.104
Difficult Task Card	58	1073	0.058

Table 3-2. Results of Pilot Study 1 for Simplified English

3.5 Results of Pilot Test 2: Native Chinese Engineering Students

From December 2003 to February 2004, we conducted a pilot test of our methodology before actually collecting data in foreign MROs in China. 40 native Chinese engineering students were recruited from the graduate student pool at the University at Buffalo. We assumed that a Chinese graduate student majoring in Engineering in the United States possessed more knowledge and had a higher ability to use the English language in general than would be typical of maintenance personnel in China. In order to decrease the gap between these two groups, we specified that student participants should have arrived in the United States less than one year ago to be eligible for this experiment. For this pilot test, we used 40 participants in a three-factor design (5 Interventions x 2 Simplified English/Not x 2 Task Cards).

For our pilot test group, there were three possible individual variables that might have affected performance: reading level score, years of learning English, and years as an Engineering major. These could have been useful covariates in the analysis of main factors by reducing the expected variability between individual participants. An inter-correlation matrix of these revealed that "Years of Learning English" was significantly correlated with the time to complete the task card comprehension questionnaire ($R = 0.498$, $p = 0.001$), and "Reading Level Score" was related to accuracy ($R = 0.34$, $p = 0.032$). We decided to consider two covariates: "Year of Learning English" and "Reading Level Score." Note that, as with Pilot Test 1, there was a negative correlation between Accuracy and Time, but here it was not statistically significant ($p = 0.091$).

We used GLM 3-factor ANOVAs on each performance variable with and without the above covariates, and found statistical significance for Time and Accuracy/Time in both cases. For Time, there was significant effect of Intervention ($F(4,20) = 7.77$, $p = 0.001$), and for Accuracy/Time there was significant effect of Task Card ($F(1,20)=5.68$, $p=0.027$). As shown in Table 3-3, the easy task card had a worse performance on all variables than did the difficult task card. The results were quite counter-intuitive, with the difficult task card having better performance than the easy one. We suspect that this may have been caused by the potential variability when two versions of each task card were translated into Mandarin. The effects of Simplified English may also have been different for the Mandarin and original versions. In fact, if the "Translation" intervention is eliminated, no terms in the ANOVA are significant.

	Accuracy, %	Time, s	Performance, %/s
Easy Task Card	66	1364	0.051
Difficult Task Card	72	1202	0.063

Table 3-3. Results of Pilot Study 2 for Simplified English

All the interventions resulted in decreased accuracy, but shorter time for completion. We did expect that these Chinese graduate students would achieve higher accuracy when comprehending a task card written in their native language. One possible explanation for this is that the aviation maintenance domain is a very specialized domain, so task cards in both English and Chinese were unfamiliar and difficult for the participants, and the advantages of the native language were somehow minimized. If we considered Performance (i.e., Accuracy/Time), all four interventions (except Glossary) resulted in better overall scores than the Control condition, and the Chinese Translation was significantly better than the Control condition at 0.068 vs. 0.500 ($T = -7.81$, $p = 0.004$). As a check on the even distribution of participants across Interventions, a one-way ANOVA of Reading Level between Interventions was conducted. As shown in Table 3-4, there were significant differences in Reading Level ($F(4,35) = 3.91$, $p< 0.01$), showing that our random assignment did not in fact produce equivalent groups.

	English Reading Level	Accuracy, %	Time, s	Performance, %/s
0 Control	10.3	75	1560	0.050
1. Glossary	11.9	73	1519	0.050
3. Tutoring	8.9	69	1264	0.056
4. Glossary & Tutoring	8.8	61	1027	0.057
2. Chinese Translation	10.6	66	1046	0.068

Table 3-4. Results of Pilot Study 2 for Interventions

Because “English Reading Level” was significantly different across Interventions, we reconsidered it as a covariate, and ran GLM 3-factor ANOVAs on each performance variable. For performance variables Time and Accuracy/Time, there was not much difference between with and without the new covariate. For Accuracy, with the covariate, the interaction between Intervention and Document English became marginally significant at ($F (4, 19) = 2.83$, $p = 0.054$).

3.6 Observations

According to our observations, most of the student participants did not utilize the interventions of glossaries, tutoring, or the combination of the above two as much as we had expected. After the experiment, the native Chinese experimenter asked them why they did not utilize the resources. The participants agreed that: “although we do not understand some words, even a sentence here and there, we are still able to answer the comprehension questionnaire; clarifying the meaning of all the details may not necessarily improve our performance, but it will take much longer to finish the task." In fact, this makes sense, as all international students who apply for graduate school in the United States need to submit their scores on the Test of English as Foreign Language (TOEFL), and the Graduate Record Examination (GRE). For non-native English speakers, in order to achieve better scores on the TOEFL and GRE-Verbal tests in a limited time, one key factor is the ability to figure out unknown words, phrases, and even sentences in context. This is a common consensus by non-native English speaking students who have gone through this process. As a result of Pilot Test 2, we have eliminated the combined Glossary and Tutoring condition from our subsequent Asian data collection.

3.7 Conclusions

The main comprehension task took less than half an hour to complete, while the other measures, such as the Reading test and the rating scales, together took another 15 minutes or so. Because many people could be tested together, we were efficient in data collection at the site, and cannot develop accurate timetables for our on-site work in China.

This experiment used a baseline condition of English documents, and then added translation (including the test form), a glossary, a bilingual coach, and a combination of these last two conditions. We used two levels of task card difficulty, each with and without Simplified English. This made a three-factor factorial experiment (Intervention x Difficulty x Simplified English), with the Reading Level score as a covariate. On the samples tested so far, the US and the UK participants obviously had the baseline intervention only, whereas the Chinese-speaking engineering students had all of the interventions. At this stage, any significant effects should be treated with caution, despite the interesting findings on Simplified English and Interventions. These pilot studies are being used for testing the methodology, training the experimenters, and providing an English-speaking baseline condition.

Chapter 4. ASIA RESULTS

A general description of the characteristics of each of the nine sites is presented in Table 4-1. The data were collected from written sources, managers and focus group discussions. A primary result of this data collection was that all of the sites in China used a mixture of English and translated Chinese documentation, while in Hong Kong and Taiwan only English documentation was used.

Area	Site #	Number of Employees	Style of Using Task Card in Maintenance
Mainland China	1	> 3,500	English-Chinese
Mainland China	2	1,520	English-Chinese
Hong Kong	3	632	English
Hong Kong	4	2,500	English
Mainland China	5	1,993	English-Chinese
Mainland China	6	769	English-Chinese
Taiwan	7	~ 300	English
Taiwan	8	1,300	English
Taiwan	9	<100	English

Table 4-1. Background information on the MROs in Asia

4.1 Demographics

For each participant we recorded their Gender, Age, Years as an AMT, Years Learning English and Reading Level as given by the Accuracy Levels Test. As it was most unlikely that these demographics would remain constant across the three areas (Mainland China, Hong Kong, Taiwan), one-way ANOVAs were conducted of each demographic, except for the categorical variable of Gender that was tested using Chi-Square. All comparisons gave significant differences by Area, as shown in Table 4-2.

	Mainland China	Hong Kong	Taiwan	UK/ USA	Test Result	Significance
Number Tested	175	25	54	15		
Percent Female	25%*	4%	4%	0	$\chi^2(2) = 15.84$	$p < 0.001$
Age	33.5*	42.9	40.5		$F(2,250) = 34.7$	$p < 0.001$
Years as AMT	8.6*	18.4*	13.6*		$F(2,250) = 21.9$	$p < 0.001$
Yr. Learning English	20.1*	35.6*	27.1*	-	$F(2,243) = 79.9$	$p < 0.001$
Reading Level	4.9*	6.6	5.8	14.1	F(2,253 = 7.9	$p < 0.001$

Table 4-2. Demographics of the three areas, with mean values and test results. Note that * signifies a mean value different from the others at p<0.05 on the *post hoc* Tukey test, or Standardized Residuals test for Chi-square.

Note that Mainland China has more females represented, in a generally younger, less experienced sample with lower English exposure and reading ability. For Years as AMT and

Years Learning English, all three areas were significantly different, with Taiwan falling between the low value of Mainland China and the high value of Hong Kong.

There were no gender differences among the demographic variables using a two-factor GLM ANOVA of Area and Gender, except for Years as AMT where females (6.1 years) were less experienced than males (11.6 years) with $F(1, 247) = 6.6$, $p = 0.011$.

4.2 Incidence Questionnaire

In addition to the evaluation of the interventions, we used a questionnaire to determine the relative incidence of the seven scenarios developed earlier. A number of measures of incidence were used, including estimates of the time since last occurrence. The first analysis was of the overall response to "Have you ever encountered an error of this type?" A two-factor GLM ANOVA (Scenario x Area) of whether or not each scenario was reported resulted in significance for Scenario $F(6, 1722) = 28.2$, $p < 0.001$; for Area $F(2, 1722) = 5.3$, $p = 0.005$, and for their interaction $F(12, 1722) = 2.7$, $p = 0.002$.

Overall, the scenarios group into three sets using the Tukey post-hoc test. The most frequent set (Scenarios 6 and 1) refer to the AMT not understanding written (6) or verbal (1) instructions. The next set of three (Scenarios 7, 2 and 3) refer to poor translation of documents (7) often from English by aircraft manufacturers for whom English is not the native language, not realizing that the AMT did not understand (2) or difficulties with regional accents (3). The least frequent set (Scenarios 5 and 4) consisted of relatively rare forms of communication, placards (5) and the PA system (4).

The three areas did not produce clear-cut results in post-hoc tests. Mainland China reported a higher incidence (42%) than Hong Kong (32%), but was not different from Taiwan (39%). Also Taiwan was not different from Hong Kong. Thus, there is a hierarchy of incidence reporting, but only the extreme values are different from each other.

The interaction for incidence of each scenario is shown in Figure 4-1 for the three areas separately. Misunderstanding translations (Scenario 7) was highest in China and lowest in Taiwan, while the opposite ordering was found for misperceived language abilities (Scenario 2) and regional accents (Scenario 3). The first of these results is perhaps reflective of exposure, as Chinese sites used translation of parts of documents, which was not a strategy in the other two areas.

When the answers to the question "When was the most recent time you encountered on errors of this type?" were tabulated, it was possible to estimate the median time since the last occurrence of each scenario. A cumulative plot of probability of occurrence against time since last occurrence for each scenario was used to perform a linear interpolation of the median. The medians are shown for each scenario in Table 4-3 with the mean percentage reported from the previous analysis. As expected, the more frequently reported scenarios are the ones with the smallest median time since previous occurrence ($r = -0.817$, $p = 0.025$).

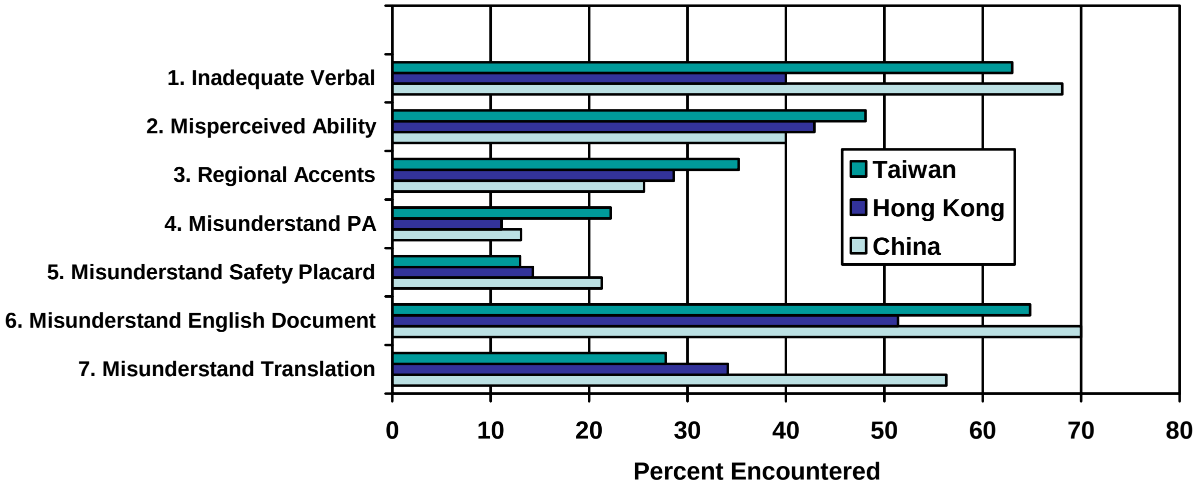


Figure 4-1. Relative frequency with which each of the seven scenarios was encountered

Scenario	Median Weeks Since Previous Occurrence	Mean Percent Reported
1. Inadequate Verbal	16.0	39.4
2. Misperceived Ability	12.9	62.1
3. Regional Accents	21.5	16.2
4. Misunderstand PA	18.9	15.5
5. Misunderstand Safety Placard	18.0	29.8
6. Misunderstand English Document	9.2	43.7
7. Misunderstand Translation	12.0	57.0

Table 4-3. Median weeks since previous occurrence and mean percent reported for each scenario

4.2.1 Error Factors

For the response to factors most associated with these scenarios, GLM ANOVA of the percentage encountering each incident by Factor was performed, with Area and Scenario as additional independent variables. All main effects and interactions except Scenario × Area were significant at $p < 0.02$ or better. Post hoc Tukey tests were performed at $p = 0.05$ to group the main effect levels of Factor. The responses divided into two groups, one group seen as highly related to the incident and one less related. These are shown in Figure 4-2 with the plotted position representing the percent reporting that factor.

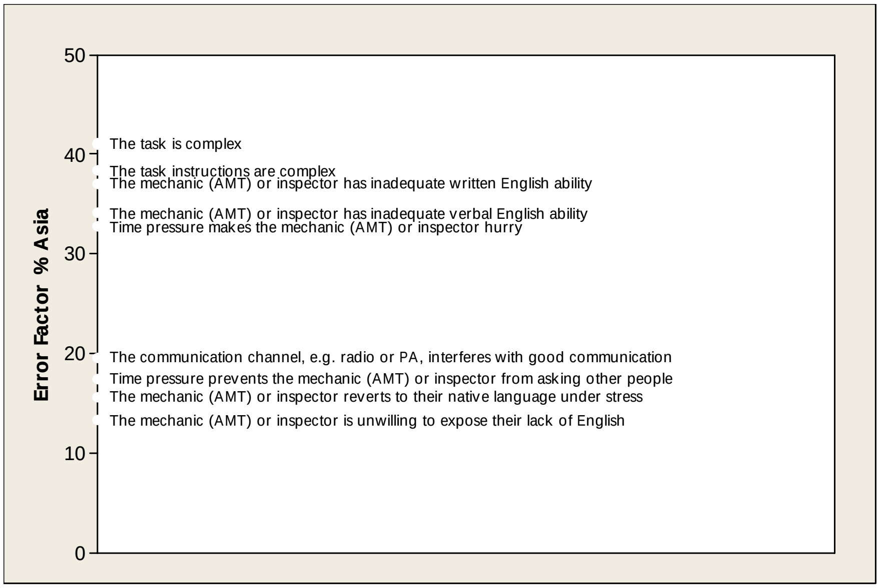


Figure 4-2. Percentage reporting each factor affecting scenario incidence

4.2.2 Prevention Factors

A similar analysis was performed for the ten factors potentially mitigating language errors. The GLM ANOVA gave significance at $p < 0.01$ for Factor, Area, and their interaction. As with causal factors, the results grouped into two:

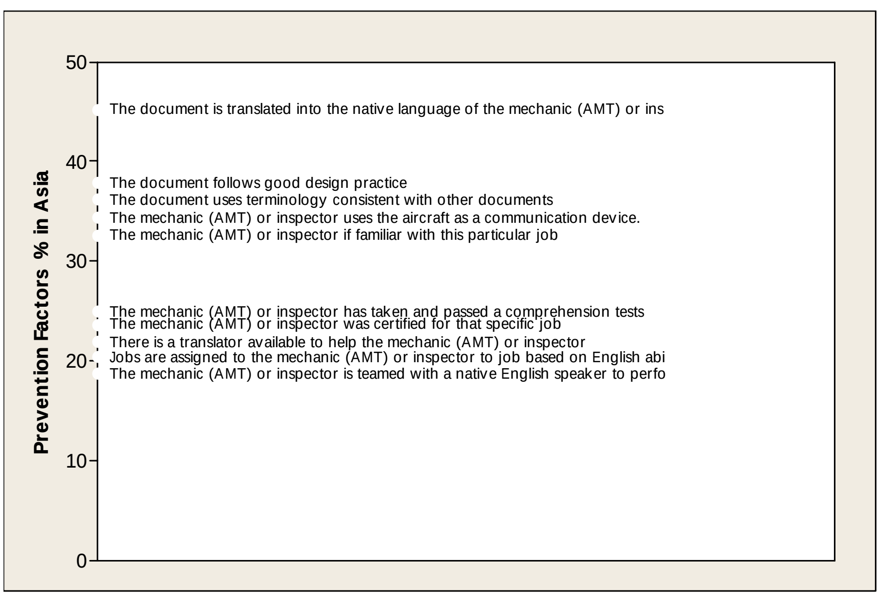


Figure 4-3. Percentage reporting each factor affecting scenario prevention

As with causal factors, the highest group included the physical changes, plus in this case job familiarity. The lowest group was mainly individual and social interventions.

4.2.3 Discovery Factors

Finally, an analysis of how errors are discovered was performed. Only Scenario, Factor, and the Factor × Area were significant (at $p < 0.02$). Again, there was a grouping of the Factors, this time into 3 groups as shown in Figure 4-4:

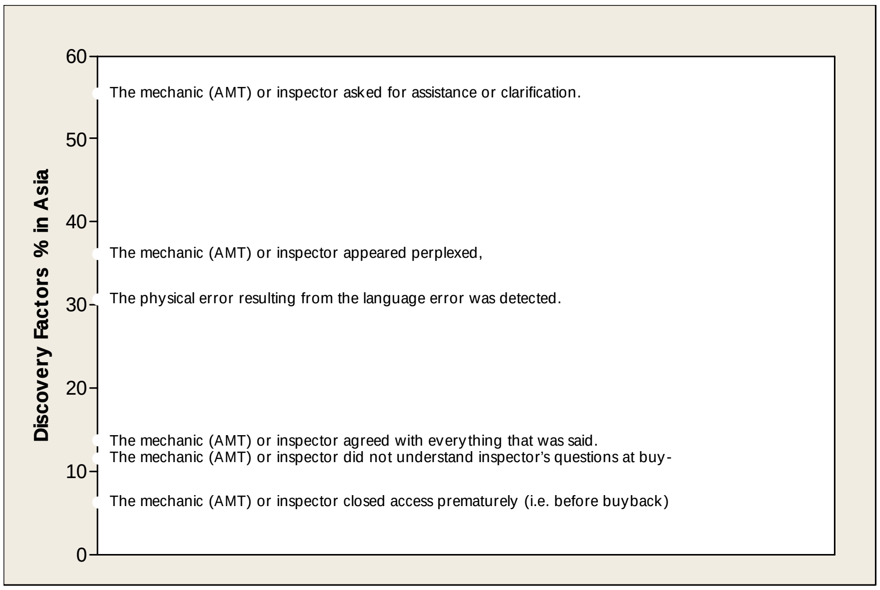


Figure 4-4. Percentage reporting each factor affecting scenario discovery

From these groupings, note that the least commonly found were either an unusual behavior, or events later in the maintenance/inspection process.

4.3 Intervention Effectiveness

This test used 254 participants from six sites in mainland China, Hong Kong and Taiwan. First, as in the pre-tests, there was a negative correlation between accuracy (fraction of correct responses) and time (overall time to complete the task) for the comprehension test ($r = -0.170$, $p = 0.007$). This was not as large as in the pre-tests, but still a significant speed/accuracy trade-off. A third measure was created by dividing Accuracy by Time to give a combined overall Performance score.

Among the demographic variables, there were inter-correlations among the measures in Years (Age, Years as AMT, Years Learning English) as would be expected, but no significant correlations of these variables with Reading Level. Another way to express this is that a Factor Analysis (using a Varimax rotation) needed only two factors to explain 86.3% of the variance in these four measures, with the first factor loading above 0.85 on all the "Years" factors and the second loading only on Reading Level. From these analyses of individual characteristics, two relatively orthogonal measures were chosen as potential covariates in the performance analyses: Reading Level and Age.

There were moderate correlations of accuracy with Years as an AMT ($r = -0.231$, $p < 0.001$) and both accuracy and time with Reading Level ($r = 0.351$, $p < 0.001$; $r = -0.250$, $p < 0.001$ respectively).

Because the Simplified English factor was not a true factor for the intervention of Chinese translation, a separate set of analyses was performed with that intervention removed. These results will be noted as similar to of different from the main analyses as each is performed. As an example, all of the correlation results in the previous paragraph were mirrored in the "no Chinese Translation" analysis.

GLM ANOVAs were performed for each measure (Accuracy, Time, Accuracy/Time) as well as Log_e(Time) because that was found to be more normally distributed than Time. The factors testes were Intervention, Area, Task Card Difficulty and Simplified English, with the two covariates of Reading Level and Age. All main effects and two-way interactions were included, but not higher order interactions due to multiple co-linearity effects. Part of that was due to the fact that the Chinese Translation intervention could not be used in Hong Kong as the participants there would only use original English documentation. The Intervention x Area interaction was dropped from the analysis because of this missing cell.

The results of the ANOVAs are summarized in Table 4-4. Note that the use of AECMA Simplified English had no significant effect on any measures. Also, no interactions among any factors reached significance, simplifying the interpretation of results. The two covariates were highly significant in all analyses, this helping to reduce the error terms and so increase the power of the other tests.

	Accuracy	**Time**	**Accuracy/Time**	**Log_e(Time)**
Intervention		$F(3, 232) = 6.1$ $P = 0.001$		$F(3, 232) = 5.9$ $P = 0.001$
Area		$F(2, 232) = 13.9$ $p < 0.001$	$F(2,232) = 13.9$ $p < 0.001$	$F(, 232) = 14.9$ $p < 0.001$
Task card		$F(1, 232) = 6.2$ $P = 0.014$		$F(1, 232) = 7.1$ $P = 0.008$
Simplified English				
Reading Level (covariate)	$F(1, 232) = 22.3$ $p < 0.001$	$F(1, 232) = 9.3$ $P = 0.003$	$F(1,232) = 18.7$ $p < 0.001$	$F(1, 232) = 7.5$ $P = 0.007$
Age (covariate)	$F(1, 232) = 17.4$ $p < 0.001$	$F(1, 232) = 11.7$ $P = 0.001$	$F(1,232) = 17.1$ $p < 0.001$	$F(1, 232) = 9.7$ $P = 0.002$

Table 4-4. Summary of ANONA results for intervention performance

To illustrate the predictive power of the covariates, Figure 4-5 shows the four plots of two aspects of performance (Accuracy, Time) against the two covariates (Reading Level, Age). While they clearly show relationships, the variance is quite high for all four plots: performance in task card comprehension is more than just good English ability and lower age.

From Table 4-4, it is obvious that most of the variation due to the four factors was seen in the speed measures (Time, Log_e(Time)) rather than accuracy. To a large extent, Accuracy remained constant across conditions. It appears that participants took as long as they needed to achieve their ultimate level of accuracy, which is a safe and conservative approach to this test.

Interventions were only different on the Time measures, although their Accuracy/Time measure approached significance at $p = 0.069$. The mean times and accuracies for the four interventions are given in Table 4-5. *Post hoc* Tukey tests showed that for times only the slowest (No intervention) and the fastest (Chinese translation) differed significantly at $p < 0.05$.

Intervention	Mean Accuracy, percent	Mean Time, s	Accuracy / Time (%/s)
1. No Intervention	73.2	1638	4.9
2. Chinese Translation	72.0	1367	5.6
3. Bilingual Glossary	73.8	1469	5.6
4. Bilingual Coach	78.2	1437	5.9

Table 4-5. Performance results for the four interventions
Shaded results not significant at $p < 0.05$

The three areas also differed on Time, but also Accuracy/Time. *Post hoc* Tukey tests at $p < 0.05$ showed that for both measures, the best performing area (Hong Kong) was significantly different from the other two (China, Taiwan). Table 4-6 shows these results.

Area	Mean Accuracy, percent	Mean Time, s	Accuracy / Time
1. China	73.0	1519	5.3
2. Hong Kong	78.2	1128*	7.3*
3. Taiwan	75.9	1506	5.5

Table 4-6. Performance comparisons by Area. Shaded results not significant at $p < 0.05$. Note that * signifies a mean value different from the others at $p < 0.05$ on the *post hoc* Tukey test

Finally, the results for the two task cards (Table 4-7) were only significant for the two speed measures, with the Easy task card being faster than the Difficult one as expected. Note that this did not happen in our pre-test with Chinese graduate students at an American university.

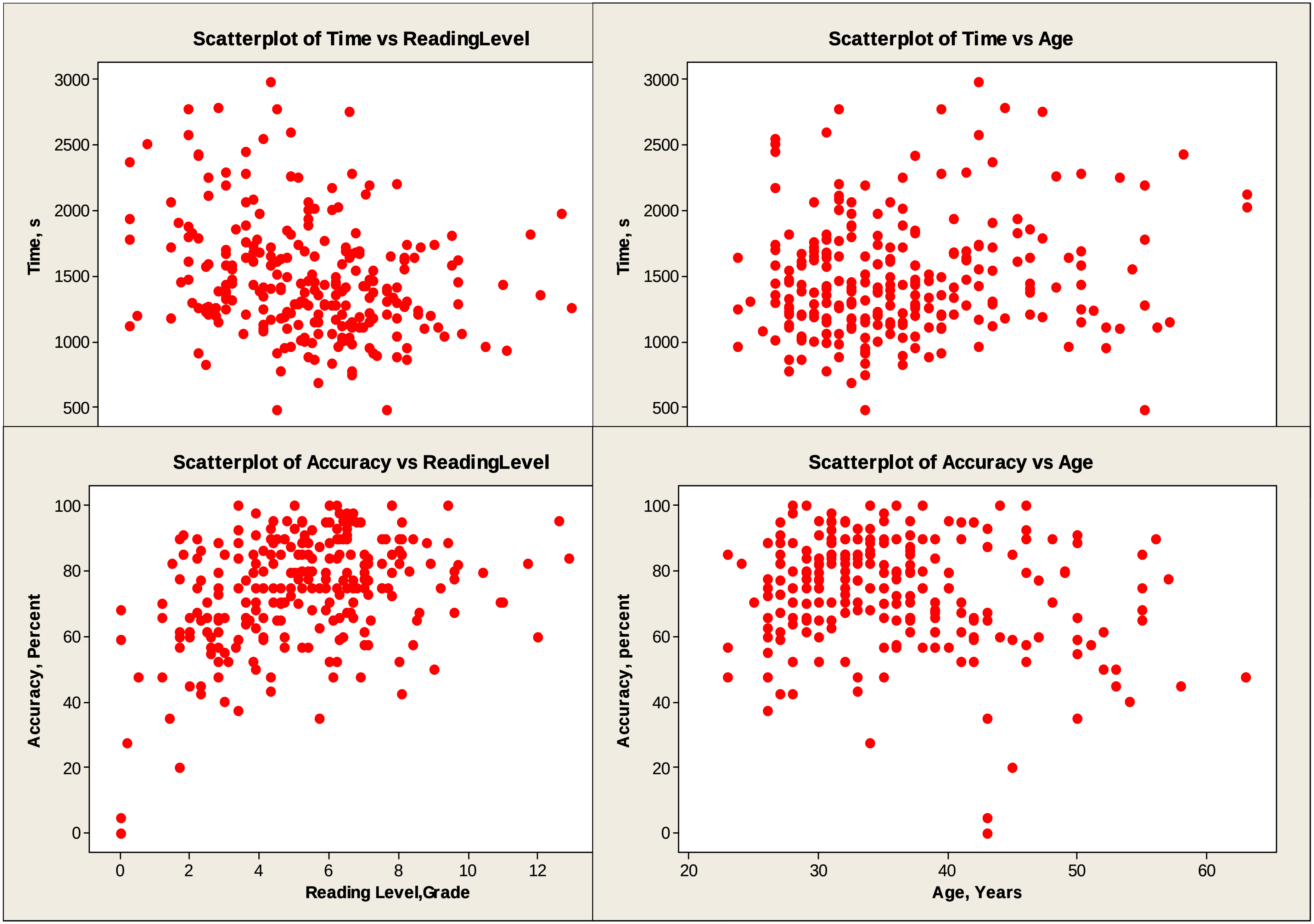


Figure 4-5. Scatter plots of the two aspects of performance (Accuracy, Time) against the two covariates (Reading Level, Age)

Task Card	Mean Accuracy, percent	Mean Time, s	Accuracy / Time
1. Easy	73.5	1373	5.9
2. Difficult	74.7	1580	5.2

Table 4-7. Performance comparisons between the two task cards. Shaded results not significant at $p < 0.05$

4.3.1 Rating scales

Identical GLM ANOVAs were performed on the fourteen rating scale values, i.e. using Reading Level and Age as covariates and Area, Task card, Simplified English and Intervention as factors. The major pattern to the results was that on 11 of the fourteen scales, Intervention was the only significant factor, with $p<0.001$ in all of these cases. For 8 of these 11 scales, the only difference in post hoc Tukey tests at $p = 0.05$ was between translation and non-translation. One of the 11 scales showed no significant contrasts while the remaining two only found translation different from the Glossary condition. In all cases, the Chinese translation was rated worse than the other interventions, perhaps reflection the participants' concerns for accuracy of translation from original English documents. Figure 4-6 compares the mean scale ratings of translation and non-translation interventions for all 15 scales, whether significant (11 scales) or not (4 scales) as noted in the caption.

The other significant results for rating scales were few:

- Task card X Simplified English for rating scale 2 (Continuity of information) $p = 0.023$
- Area for rating scale 4 (Chance of missing information) $p < 0.001$, and also for rating scale 12 (Compatibility with supplementary information) $p < 0.001$. For both, the only difference was that Taiwan rated significantly lower than the other two areas.
- Age was a significant covariate for Rating Scale 13 (Amount of graphical information)

4.4 Focus Groups

The major characteristics of each site can be found in Table 4-1. Supplementing this information, a native English-peaking moderator and an English/Chinese bilingual assistant facilitated the 11 focus groups held across all sites. Focus groups were encouraged to use the language they felt most comfortable with during the discussion: Chinese, English, or both. Each session lasted about 30-45 minutes and was audio taped with. The main points are summarized below from combined notes and transcripts.

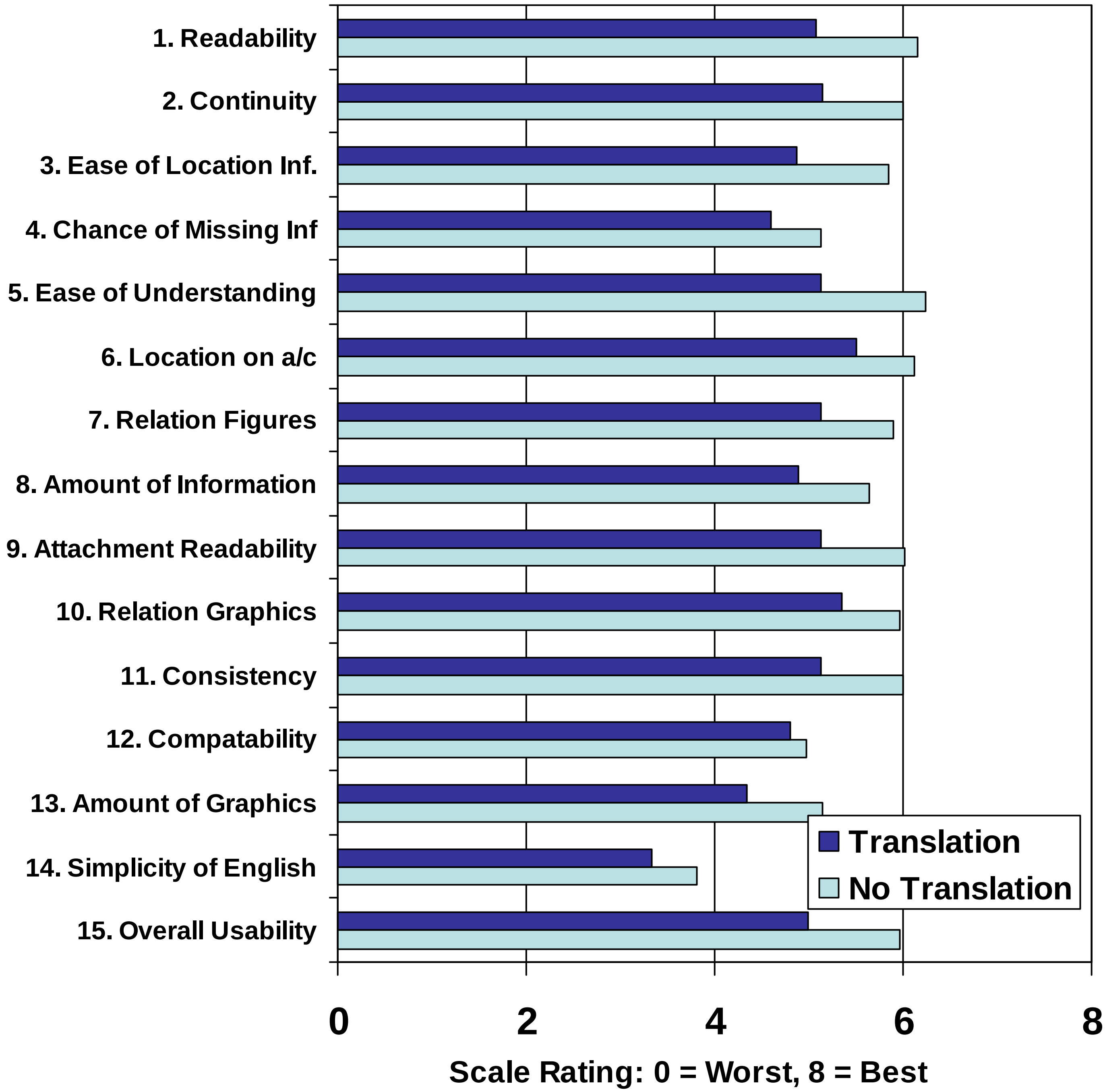


Figure 4-6. Differences between Chinese translation and the average of all non-translation groups on the fifteen rating scales

4.4.1 Current Practice

1. Written Communication to AMTs

- In China, task cards are bilingual, while Non-Routine Repair forms (NRRs) can be either in Chinese or English; [however] companies encourage English. In contrast, task cards and NRRs are in English for both Hong Kong and Taiwan.

- Maintenance manuals are in English for all three areas. The focus groups agreed that Aircraft Maintenance Manual's English is relatively simple (e.g., simple grammar and sentence structure, short sentences).

- Focus groups complained about difficulty in comprehending aspects of English:
  - Long sentences in FARs, JAR, especially those documents related to legal interpretations. In this case, even the Chinese translation is difficult to understand. Even the original English documents can be ambiguous, which results in misunderstanding. In particular, English originals may not be detailed enough, with many steps omitted by the editors of task cards.
  - Multiple meaning for an English word (especially abbreviation). Abbreviation is a general problem: many questions about English meanings are questions about abbreviations. Different manufactures use also different words or phrases to describe the same thing in their manuals.

- The same English words may be translated into different words in China, Hong Kong, and Taiwan.

2. Writing communication from AMTs

- The shift hand over document in often in English and Chinese, or in Chinese with technical words left in English

- NRRs are written in Chinese by the mechanics, and then translated by engineer and manager because the international customers require English to be used in NRRs.

- There is a distinct Chinese style of English that can be understood by fellow Chinese colleagues but not by non-Chinese colleagues and manufacturer's representatives. Considerable management effort is spent on rewriting English written by the employees.

3. Verbal communication

- In China:

  - Most engineers, QA personnel, and leads/foremen speak English. At one site in China, three languages were used in the production meeting every morning.

  - The technicians' oral English ability is often poor.

  - The level of English and Technical English is good in the young generation of aircraft maintenance trainees, but their oral English is still poor.

- At sites in Hong Kong, there is a barrier between the “Mandarin” and “Cantonese” sub-languages e.g., at the maintenance control center.

- Local “nick names” are used in daily work, e.g., “turtle shell.” Everybody in the shop knows what it is, but other people do not necessarily know, and cannot write it down.

4. Company policy

- Focus groups agreed that English is aviation language, which is a fact that nobody can change. There should not be localized variations. All personnel should emphasize reinforcing standardization inside the company. The focus groups believed that it perhaps the company’s fault that it allows two languages co-exist at work place. MROs must be more aware of international standards to stay competitive, e.g., comparing themselves with other maintenance bases that only use English job cards.

- In reality, MROs will continue work with manufacturers that use very different English: French English, Brazilian English, and American English. These manufacturers may use very different names for the same thing. MROs will also continue work with manufacture representatives who have different language backgrounds (e.g., American English vs. British English).

- New technology has brought changes to the aviation vocabulary. New words cannot be found in the dictionary, even onsite manufacturer representatives were not sure about them.

4.4.2 Intervention Methods

1. Better design of documentation

- Translation of task cards was the option used at the sites in China, including initial translation, auditing, and second auditing. Feedback forms to report problems identified onsite with the translated task cards were available, as they are in all organizations using task cards. Focus groups agreed that translation might not be the ideal solution, because:

    - Translations are currently done by college Chinese graduates who are English majors. They have relatively shallow comprehension of aircraft, and short (or even no) working experience on the aircraft. The translation is often based on the obvious non-aviation meaning of the English without in the context of the aircraft. The mechanics find the translations can be confusing, awkward, and even strange.

    - The sheer amount of translation/auditing involves expenses of staff plus an overhead loss and about 30-40 % of the total maintenance time. These all increase maintenance services cost and make the MROs less competitive.

    - The translation/technical writing/editing group has many personnel, each with their own styles in choosing words and structure sentences, e.g., calling a part several different names, which can be confusing to the Chinese mechanics.

- Provision of both English and Chinese translated versions can help but it is not the final solution—simply because it is impossible to translate everything. For example, there are frequent modifications from the manufacture.

    - Sometimes reading originals in English is easier than using Chinese translation, especially where the Chinese meaning and English meaning sometimes don't match very well. Occasionally, translation makes it worse, e.g., "on/off" "close/open" can be translated into exactly the same Chinese words.

    - Translations of technical references, operation procedures/materials/tools can't be exactly perfect. There must be English originals available.

- Most MROs have dedicated special focused effort (e.g., company training center, language committee) to develop language references such as:

    - Abbreviation/acronym dictionaries.
    - Glossaries, which were developed by "data mining" for most used Chinese words in the Maintenance Manual.
    - A "Pocket book" consisting of a Chinese-English/English-Chinese dictionary. Most mechanics carry a well-worn copy of this pocket book

- Focus groups demanded that original English documents use standardization and Simplified English in order to:
    - Be able to use translation software.
    - Clarify the confusion caused by non-native English speaker's lack backgrounds of words.
    - This is especially true for regulations, technical stuff, e.g., get rid of the double negative, which can be confusing rather than emphasizing.)

- Mechanics do appreciate diagrams. - Increase numbers of illustrations and diagrams (especially emphasize different angles and positions)

- All would prefer manufactures to provide reference links in its maintenance manual CD-ROMs to reference other documents.

2. Better Education, Training and Language assistance

- English ability criteria have been used to hire and evaluate performance. English classes have become a part of the curriculum to train apprentices. Some technical classes will be taught in English in the near future. Apprentices are required to pass specific English tests to graduate, and more tests to become certified or promoted. Performance evaluation should always include English. Require certification of English ability integrating with technical/management types of certifications. A small number of employees are selected to study English in local universities every year.

- Engineers are typically on call 24/7 for help with English on project-by-project basis. People are good at going to supporting engineers for help. However, an engineer often works with many mechanics on the same shift. Mechanics consult engineers for trouble shooting, e.g., checking the Chinese translation and English originals.

Finally, the focus groups have confirmed that there are incidents caused by language barriers. Some examples are:

Case #1: One MRO had an incident caused by "language" resulting in engine damage in 2001. The English word "Clean" has two meanings: 1) get rid of paint, e.g strip, and 2) use cleaner to clean. The correct interpretation should be "get rid of paint; strip" in this context. However, The mechanic did not understand, and performed cleaning by "use cleaner to clean," which resulted wires being burned from the cleaning fluid.

Case #2: On a test procedure in a manual the Chinese translation did not correctly point out that the "115-160 voltage" should be switched rather than adjusted continuously. Damage to the aircraft resulted.

Case #3: Slipping Ladder for emergency door: Different people have written the descriptions of the emergency door in different places in the Maintenance Manual. They used different words in different places of the task card to mean the same thing. The mechanics could not tie the safe-wire the way it was illustrated in the manual. In the end, they had to discuss the problem with the manufacture and follow their faxed instruction and illustrations. Due to time difference and language barriers, the discussion lasted over 3 days, which prolonged the maintenance process.

Chapter 5. LATIN AMERICA RESULTS

The data were collected from written sources, managers and focus group discussions. Table 5-1 shows the sites and countries with characteristics of each site. A primary result of this data collection was that all of the sites in Latin America used only English documentation, although translation was attempted several years ago in Mexico in accordance with government policy.

Area	Site #	Number of Employees	Style of Using Task Cards in Maintenance
Mexico	1	890	English
Mexico	2	782	English
Mexico	3	1200	English
Mexico	4	70	English
Puerto Rico	5	42	English
Colombia	6	1159	English
Argentina	7	249	English

Table 5-1. Background information on the MROs in Latin America

5.1 Demographics

For each participant we recorded their Gender, Age, Years as an AMT, Years Learning English and Reading Level as given by the Accuracy Levels Test. One-way ANOVAs were conducted of each demographic, except for the categorical variable of Gender that was tested using Chi-Square. All comparisons gave significant differences by Country, as shown in Table 5-2.

	Mexico	PR	Colombia	Argentina	Test Result	Significance
Number Tested	250	25	86	141		
Percent Female	4.0	0	7.0	2.9	$\chi 2$ (3) = 4.2	ns
Age	37.3*	35.6	34.6	31.8*	F(3,494) = 9.69	$p < 0.001$
Years as AMT	12.1	11.2	10.1	10.5	F(3,468) = 1.44	ns
Yr. Learning English	3.5	9.5*	3.3	4.4	F(3,363) = 10.38	$p < 0.001$
Reading Level	5.6	10.0*	4.9	4.8	F(3,498) = 27.63	$p < 0.001$

Table 5-2. Demographics of the four countries, with mean values and test results. Note that * signifies a mean value different from the others at p<0.05 on the *post hoc* Tukey test.

Throughout Latin America there were very few females in the aircraft maintenance population tested, and no significant differences between countries. Age distributions were lowest in Argentina and highest in Mexico. For Years as AMT there were no country differences. Years Learning English, and Reading Level both showed that Puerto Rico was much more experienced and fluent than the other countries, as would be expected by its US Commonwealth status.

5.2 Incidence Questionnaire

In addition to the evaluation of the interventions, we used a questionnaire to determine the relative incidence of the seven scenarios developed earlier. A number of measures of incidence were used, including estimates of the time since last occurrence. The first analysis was of the overall response to "Have you ever encountered an error of this type?" A two-factor GLM ANOVA (Scenario x Country) of whether or not each scenario was reported resulted in significance for Scenario $F(6, 18) = 18.3$, $p < 0.001$ and for Country $F(2, 18) = 5.4$, $p = 0.008$.

The incidence of each scenario is shown in Figure 5-1 for the four countries separately. Overall, misunderstanding the English documents was the highest frequency of the seven scenarios. Misunderstanding translations (Scenario 7) was highest in Mexico and lowest (zero) in PR. As in Asia, the three middle scenarios (3, 4 and 5) were much the lowest for all countries.

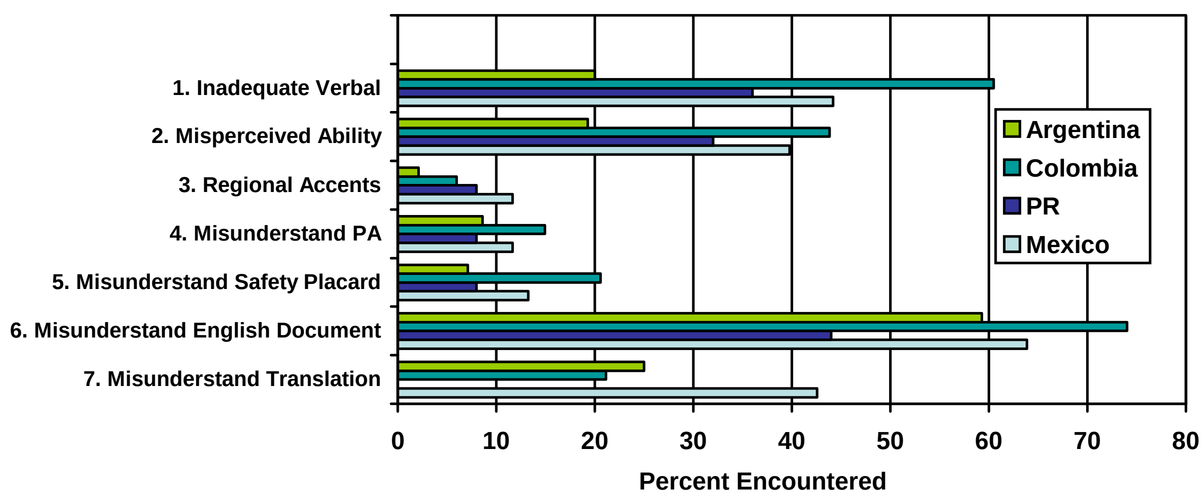


Figure 5-1. Relative frequency with which each of the seven scenarios was encountered

When the answers to the question "When was the most recent time you encountered on errors of this type?" were tabulated, it was possible to estimate the median time since the last occurrence of each scenario. A cumulative plot of probability of occurrence against time since last occurrence for each scenario was used to perform a linear interpolation of the median. The medians are shown for each scenario in Table 5-3 with the mean percentage reported from the previous analysis. Unlike Asia, there was no significant correlation between the two numeric columns of Table 5-2 ($r = -0.464$, $p = 0.294$).

Scenario	Median Weeks Since Previous Occurrence	Mean Percent Reported
1. Inadequate Verbal	6.1	40.2
2. Misperceived Ability	7.4	33.7
3. Regional Accents	96.1	6.9
4. Misunderstand PA	19.4	10.8
5. Misunderstand Safety Placard	47.5	12.2
6. Misunderstand English Document	31.6	60.3
7. Misunderstand Translation	11.0	22.2

Table 5-3. Median Weeks Since Previous Occurrence and Mean Percent Reported for each scenario

5.2.1 Error Factors

For the response to factors most associated with these scenarios, GLM ANOVA of the percentage encountering each incident by Factor was performed, with Country and Scenario as additional independent variables. All main effects and interactions except Scenario × Factor were significant at $p < 0.01$ or better. Rather than present all of these significant effects in detail, we summarize the percents reporting each factor as Figure 5- 2.

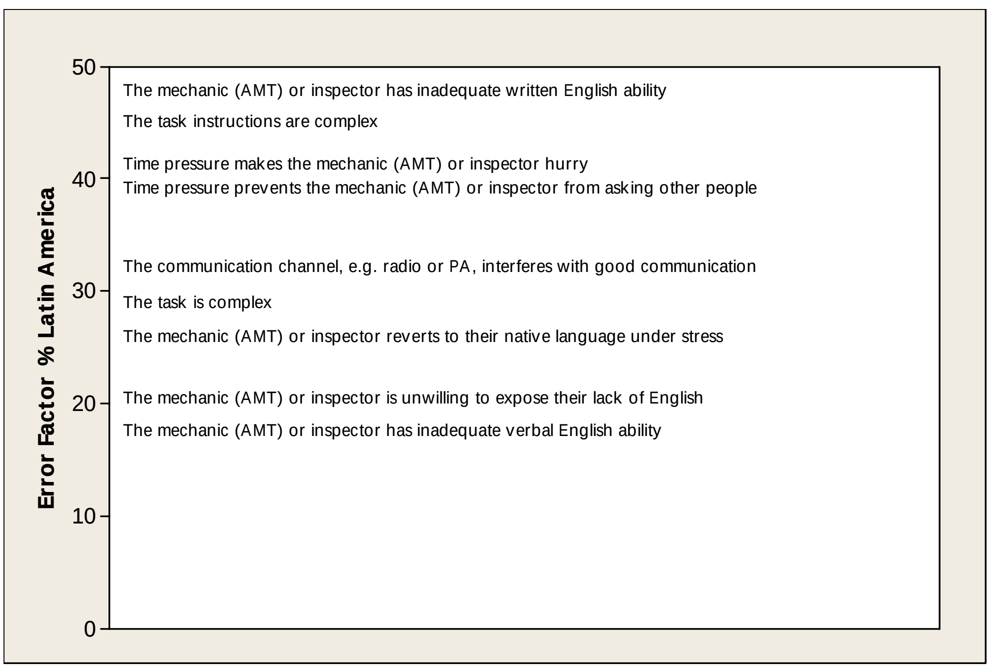


Figure 5-2. Percentage reporting each factor affecting scenario incidence

5.2.2 Prevention Factors

A similar analysis was performed for the ten factors potentially mitigating language errors. The GLM ANOVA gave significance at $p < 0.001$ for all factors and interactions except Scenario × Factor. The main effect of Factor is shown as Figure 5- 3.

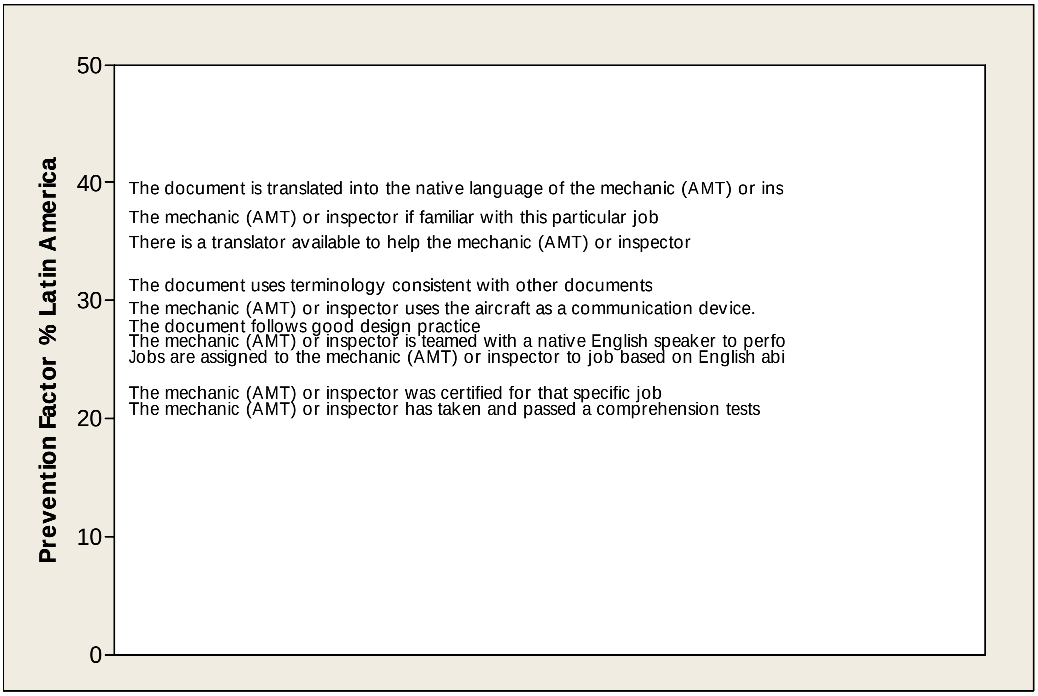


Figure 5-3. Percentage reporting each factor affecting scenario prevention

5.2.3 Discovery Factors

Finally, an analysis of how errors are discovered was performed. The GLM ANOVA gave significance at $p < 0.05$ for all factors and interactions except Scenario × Factor. The main effect of Factor is shown as Figure 5-4.

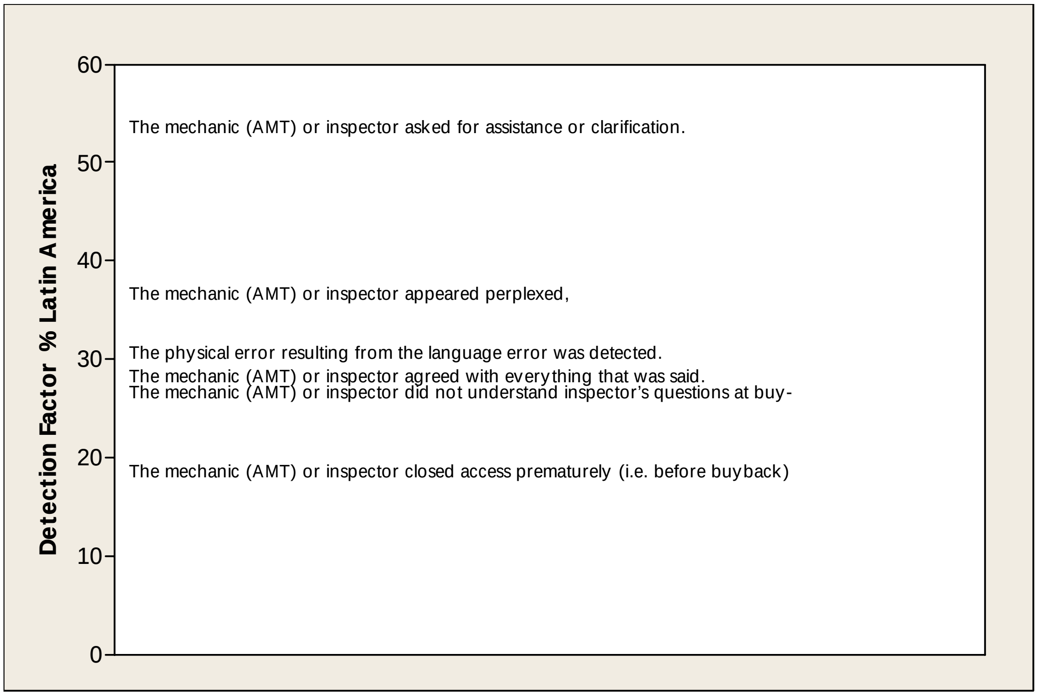


Figure 5-4. Percentage reporting each factor affecting scenario discovery

5.3 Intervention Effectiveness

This test used 502 participants from seven sites in four countries of Latin America. First, as in the pre-tests, there was a negative correlation between accuracy (fraction of correct responses) and time (overall time to complete the task) for the comprehension test ($r = -0.293$, $p < 0.001$). This was a significant speed/accuracy trade off, larger than found in Asia. A third measure was created by dividing Accuracy by Time to give a combined overall Performance score.

Among the demographic variables, there were inter-correlations between Years of Age and Years as an AMT as would be expected, and also a significant correlation of Years Learning English with Reading Level. This latter shows that years spent in English language study does indeed pay off in better reading performance. Another way to express this is that a Factor Analysis (using a Varimax rotation) needed only two factors to explain 79.4% of the variance in these four measures, with the first factor loading Years of Age and Years as AMT and the second loading on Years Learning English and Reading Level. From these analyses of individual characteristics, two relatively orthogonal measures were chosen as potential covariates in the performance analyses: Reading Level and Age.

There were moderate correlations of both accuracy and time with Age ($r = -0.207$, $p < 0.001$, $r = -0.121$, $p = 0.007$) and higher correlations of accuracy and time with Reading Level ($r = 0.551$, $p < 0.001$; $r = -0.394$, $p < 0.001$, respectively).

GLM ANOVAs were performed for each measure (Accuracy, Time, Accuracy/Time) as well as Log_e(Time) because that was found to be more normally distributed than Time. The factors tested were Intervention, Country, Task card Difficulty and Simplified English, with the two covariates of Reading Level and Age. All main effects and two-way interactions were included except Intervention x Simplified English as the Spanish translations did not differ between Simplified English and non-Simplified English. Higher order interactions were not included due to multiple co-linearity effects.

The results of the ANOVAs are summarized in Table 5-4. Note that the use of AECMA Simplified English had no significant effect on any measures, but interacted with Task card on a number of measures. The only other interaction among any factors that reached significance was Task card x Intervention for two measures involving accuracy. The two covariates were highly significant in all analyses, this helping to reduce the error terms and so increase the power of the other tests.

	Accuracy	**Time**	**Log$_e$(Time)**	**Accuracy/Time**
Intervention				
Country	$F(3,463) = 9.17$ $p < 0.001$	$F(3,461) = 5.1$ $P = 0.002$	$F(3,461) = 4.5$ $P = 0.004$	$F(3,461) = 10.5$ $p < 0.001$
Task card		$F(1, 461) = 5.2$ $P = 0.023$	$F(1,461) = 6.0$ $P = 0.015$	
Task card x Intervention	$F(4,463) = 2.4$ $P = 0.021$			$F(5,461) = 2.7$ $P = 0.032$
Task card x Simplified English		$F(1,461) = 5.5$ $P = 0.020$	$F(1,461) = 4.7$ $P = 0.031$	$F(1,461) = 5.6$ $P = 0.018$
Reading Level (covariate)	$F(1, 463) = 258$ $p < 0.001$	$F(1, 461) = 106$ $P < 0.001$	$F(1,461) = 121$ $p < 0.001$	$F(1, 461) = 367$ $p < 0.001$
Age (covariate)	$F(1, 463) = 22.0$ $p < 0.001$	$F(1, 461) = 7.0$ $P < 0.001$	$F(1,461) = 8.5$ $P = 0.004$	$F(1, 461) = 29.2$ $p < 0.001$

Table 5-4. Summary of ANONA results for intervention performance

To illustrate the predictive power of the covariates, Figure 5-5 shows the four plots of two aspects of performance (Accuracy, Time) against the two covariates (Reading Level, Age). While they clearly show relationships, the variance is quite high for all four plots: performance in task card comprehension is more than just good English ability and lower age.

In contrast to the Asia results the variation due to the four factors was seen in both accuracy and speed measures (Time, Log$_e$(Time)). The participants varied both speed and accuracy in response to changing task card and intervention, and they also differed by Country. There were no main effects of Intervention. The effect of Task card X Intervention is shown in Figure 5- 6 for Accuracy. For the Easy Task card, Intervention makes no difference, but for the Difficult Task card, the two translation conditions result in higher accuracy than the other conditions, raising accuracy from a mean of 59% to a mean of 75%. This reduces error rate from 41% to 25%, and is a worthwhile improvement in comprehension.

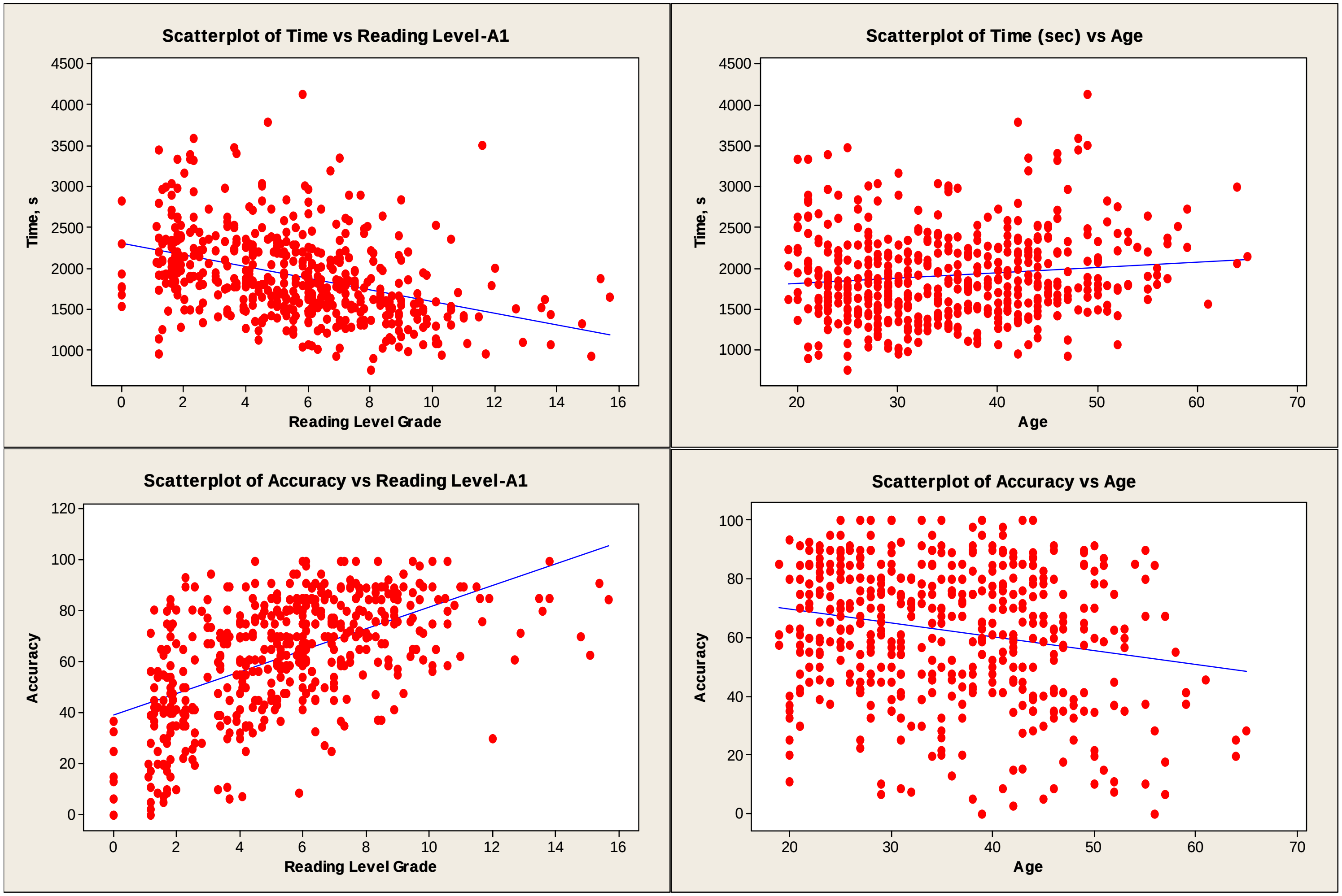


Figure 5-5. Scatter plots of the two aspects of performance (Accuracy, Time) against the two covariates (Reading Level)

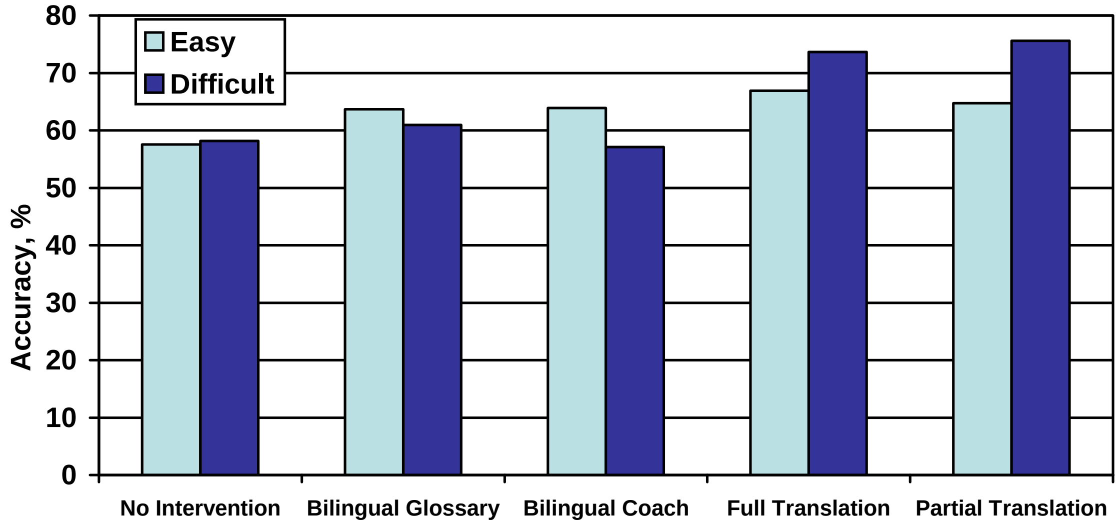


Figure 5-6. Significant interaction between Intervention and Task Card Difficulty (see text)

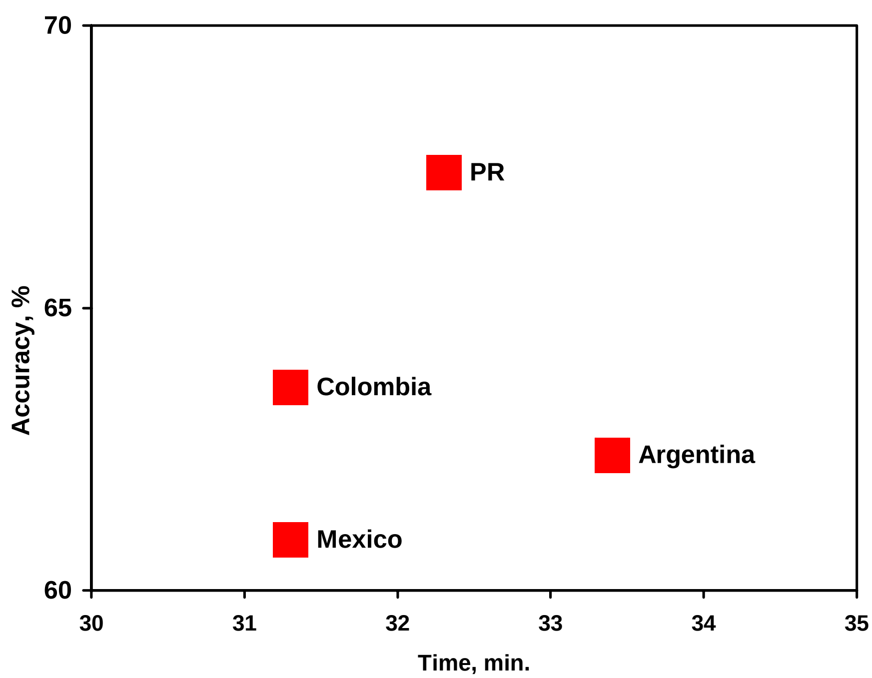


Figure 5-7. Speed and Accuracy across the three countries

The four countries differed on both Accuracy and Time, as shown in Figure 5- 8 by the raw mean results. A post-hoc Tukey test at p = 0.95 showed that Puerto Rico differed from the other three countries on accuracy, being more accurate by about 5 points than the other three. For Time, the raw means are plotted on Figure 5- 7, and Puerto Rico has a time in the middle of the countries. However, using GLM adjusted means, Puerto Rico was significantly slower than the other countries using the same Tukey test. A country where bilingualism is the norm, and where the English education and reading levels are higher, proved more accurate but slower than other countries in the region.

Task card difficulty interacted with Simplified English for Time and Accuracy/Time with the latter shown in Figure 5-8. The expected effects of the Easy Task card having better performance than the Difficult one, and Simplified English out-performing Non-Simplified English are clearly shown. The interaction, although significant, is not large in practical terms.

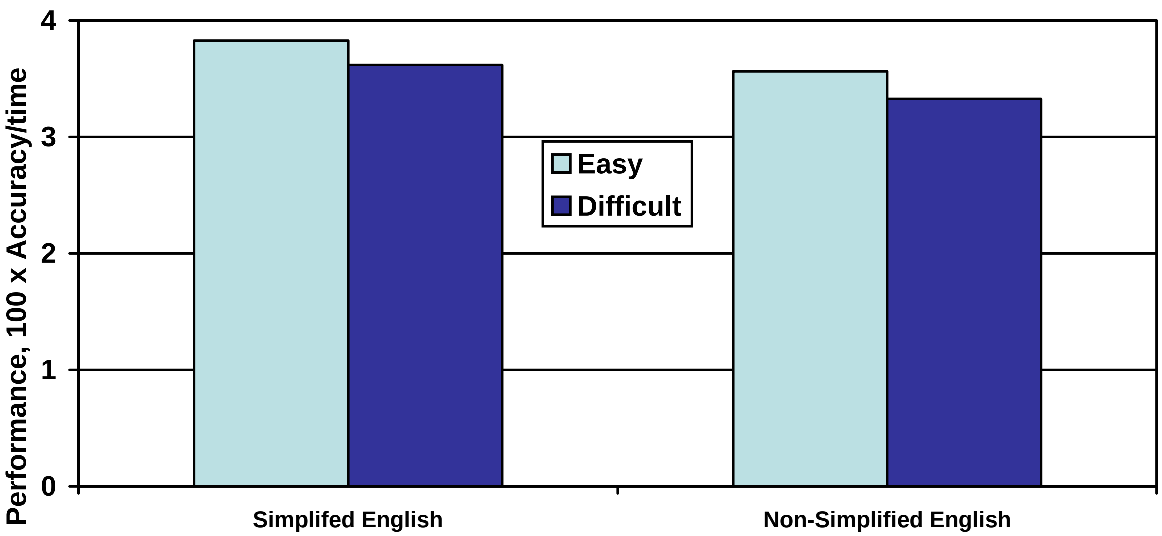


Figure 5-8. Interaction of Task Card and Simplified English for Accuracy /Time

5.4.2 Rating scales

Identical GLM ANOVAs were performed on the fourteen rating scale values, i.e. using Reading Level and Age as covariates and Country, Task Card, Simplified English and Intervention as factors. There were no significant main effects or interactions involving intervention. Figure 5-9 compares the mean scale ratings for all 15 scales, showing that the task cards were generally highly rated. Significant effects (at $p<0.05$) of Country were found for scales 1, 5, 9, 11, 14 and 15, with Argentina having considerably lower ratings than the other three Countries on all of these scales. Task card was significant for “13. Amount of Graphics” ($F(1, 439) = 22.75$, p<0.001) with the Easy task card rated much better (5.3 vs. 4.1) than the Difficult task card. Covariates were significant for Age on scales 1, 5 and 13 and Reading Level on scales 2, 3, 4, 5, 7 and 14. For Age ratings of Readability and Chance of Missing Information decreased with age, while for Amount of Graphics, ratings increased with age. Correlations were quite low, all less than 0.2. Reading Level correlations were also low, less than 0.25, and all positive except for simplicity of English Used which was judged lower for participants with higher reading levels.

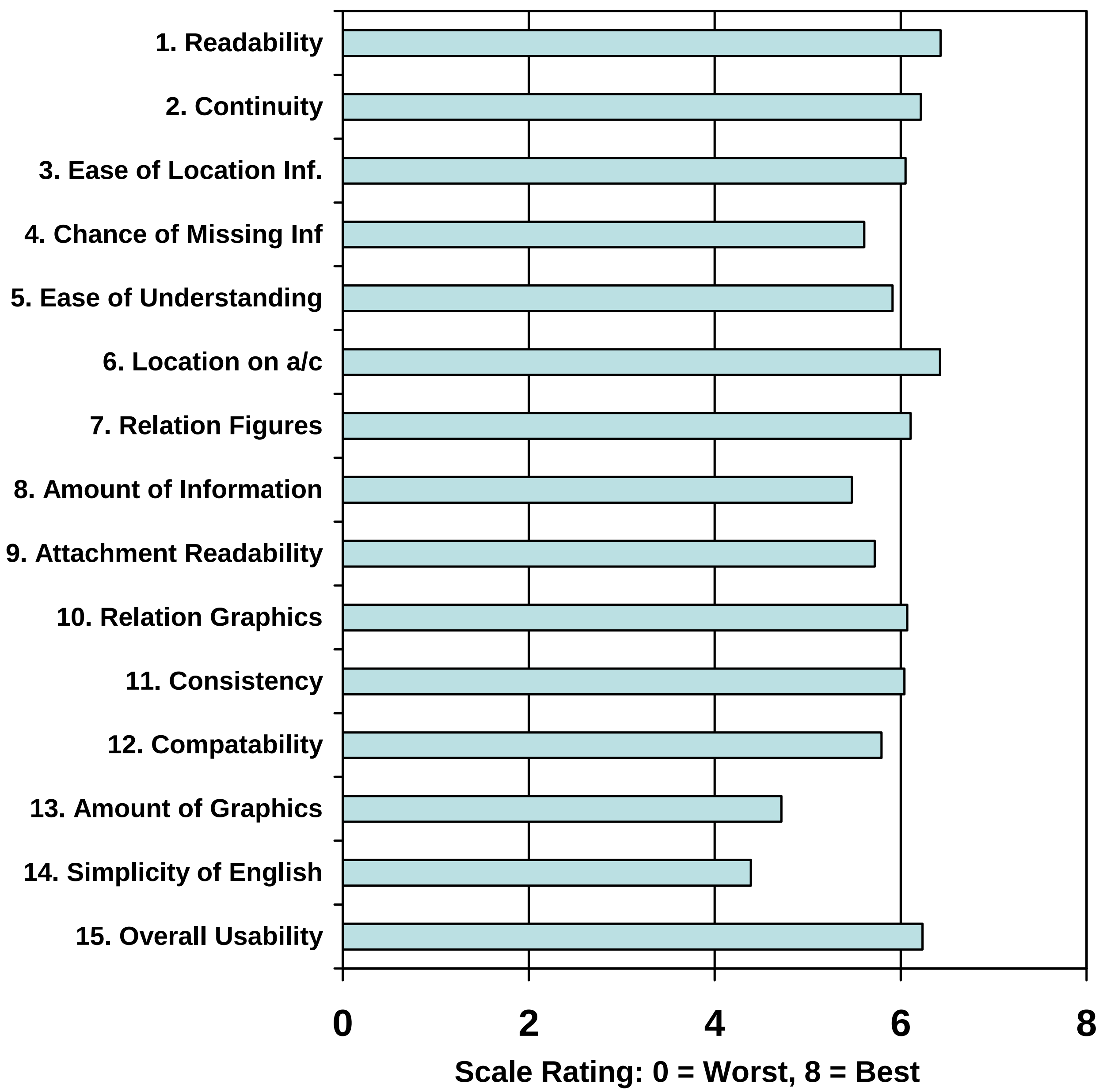


Figure 5-9. Averages of the fifteen rating scales

5.4 Focus Groups

5.4.1 Background information

In Mexico, during the 1960s and1970s, a law from the Direccion General Aeronautica Civil (DGAC) enforced Spanish as the official language used on all the written materials provided to workers. Therefore, repair stations needed to translate all manuals and work cards to Spanish. This rule increased the workload (and overtime) for bilingual engineers and the workforce was augmented with many translators to try to keep all the technical manuals updated. However, it was difficult to keep the pace of updating new versions and for many different aircraft fleets. Many times when the translated information was ready to be used by the AMTs, it was already obsolete. Due to the high cost and inefficiency of the translation intervention, the government allowed repair stations to use the original manufacturers' manuals. A mutual verbal agreement was obtained whereby manufacturers provide clearer and simple maintenance manuals (use of simple commands) and the repair stations assure that inspectors, supervisors and engineers are competent in English. When hiring, repair stations test new personnel in English skills and employees who improve their English proficiency are candidates for promotions.

Five of the seven sites used in Latin America were airline-based, although they all performed Part 145 services under FAA license. The exceptions were a helicopter overhaul facility in Mexico and a Part 145 operation in Puerto Rico that worked on some military helicopters as well as civil aircraft. All had quite extensive capabilities, with most having component repair shops and NDI departments. At least one focus group was conducted at each site, typically using Spanish and English as appropriate. Work practices were discussed, under the same headings used at the Asia sites. Where there were differences between sites or countries, these are brought out separately. In Mexico there is generally a low turnover among AMTs.

5.4.2 Written communications to AMTs

- *Manufacturer's Maintenance Manual:* Typically unchanged from manufacturer, thus in English. A few are translated into Spanish. Note that for companies using French aircraft and engines, French is also used for some manuals and written materials.
- *General Maintenance Manual:* Both English and Spanish versions are typically available for AMTs. Some of the Spanish manuals originated in Spain. In some companies only Spanish is available.
- *Task cards:* For most maintenance and inspection tasks, the task cards are not translated into Spanish to reduce potential errors. However, in component shops, the manuals are English but the task cards are in Spanish.
- *Engineering Orders*: Most of the engineering work orders are bilingual (Spanish / English), although some companies use Spanish only. A few engineering work orders come in Spanish directly from the Airlines.
- *Non-Routine Repair:* NRR work documents are in Spanish only in most companies, but both languages English and Spanish simultaneously in a few. In one company, work orders from inspectors are sent to supervisors who then write in Spanish to AMTs.
- *Shift change Documents*: Shift change forms from supervisors to AMTs are in Spanish.

- *Contracts:* All contracts are in English, except when outsourcing services to other local companies where Spanish is used.
- *Log Books*: For international flights English is used for all written documents. For domestic airlines log books are in Spanish.
- *Audits*: All audits from FAA and JAA and aircrafts' owners are performed in English. Audits for local authorities are in Spanish.
- *Warnings and safety instructions* are bilingual.

5.4.3 Writing communications from AMTs

All communications are in Spanish. Internal documents between departments are in Spanish. In Puerto Rico, both Spanish and English are used. Some technical personnel responsible for FAA audits write all documentation in English.

5.4.4 Verbal Communication:

In general, verbal communications (informal and formal meetings) are in Spanish with the exception of audits performed by international entities. FAA 145 does an annual inspection. FAA 145 inspectors select randomly employees and ask them about specific procedures in English. At the site maintaining French-sourced aircraft, engineers and managers also communicate in French with company headquarters. All emergency announcements are bilingual. In Puerto Rico, inside the company the communication is mostly in Spanish; however, some employees only speak English.

5.4.5 Common Errors

There are five types of situations that can potentially create an error:

1. AMTs do not understand the meaning of some new technical words. For these cases, the AMT asks his/her immediate supervisor or uses a technical dictionary.
2. AMTs do not understand the task cards because the procedure is too complex, for example older MDD models. Supervisors explain the complex paragraphs or consult with engineering to do the tasks.
3. AMTs have difficulties with words such as "replace" because it can be interpreted as either installing a new piece or fixing the old piece and installing it again.
4. The information in the tasks cards is incomplete. The engineering department contacts the manufacturers to ask for more documents to clarify the issues addressed by the AMTs.
5. Translation errors from French to English, French to Spanish or from English to Spanish.

5.4.5 Intervention Methods

Better design of documentation:

1. Manuals can be written to use simple command words and steps to do the tasks. An example is redefinitions of defects in inspection procedures for "crack" and "dent" where photographs are used as part of the definition.

2. AMTs can use a technical aeronautical dictionary (English-Spanish), either carried with them or located in the managers' office. One technical dictionary was developed by the students at the local aeronautical university. Another dictionary was made in Spain.
3. If AMTs are spending too much time (1-2 hours) reading a task card, supervisors help them either by translating to Spanish or by providing on the job training to clarify the procedure.
4. Partial translation may be used. Many technical words are used directly in English, therefore a combination of English and Spanish is used for communication between the technical personnel.
5. Companies periodically receive visits of Technical representatives from airframe and engine OEM's who clarify any questions to engineers related with the English used on manuals and tasks cards.
6. The internet is used to look for the meaning new words (e.g. www.dictionary.com).
7. In one company, the quality assurance department prepares specific instruction manuals for AMTs. These manuals include photos of step by step how to do the procedures. The steps are written in Spanish.

Training:

1. Multiple levels of English courses are provided to the technical personnel, either daily, weekly or in short courses. The courses focus both on grammar, e.g. writing technical reports and translation, and on interpretation of English documentation. Training may be voluntary or compulsory, and may be on company time or AMT's own time.
2. Experienced mechanics and supervisors provide on the job training (OJT) to the new mechanics. They give overviews of the tasks that need to be performed.
3. AMTs receive training and testing on specific systems during 2-3 days. Workshops using manufacturer manuals are provided. Teachers typically use English task cards and manuals in their classes, but the technical courses about aircraft maintenance are typically taught in Spanish.
4. Internal advice letters are used to illustrate repetitive errors and how to prevent them (e.g. warning signs, engineering and light maintenance tips).
5. For new jobs, a series of briefings are prepared for engineering to explain the tasks to supervisors. Supervisors are responsible to explain in detail the new tasks to AMTs.
6. Managers are often responsible to assign people who needed improve their English skills to specific training courses.
7. For technical personnel, English courses are also offered. In companies using French, French courses are offered regularly for secretaries and technicians.
8. Puerto Rico is a bilingual country, therefore, English is taught in high school and college levels.
9. Some companies have a full time English teacher available to AMTS to help them improve their English skills.

Language assistance:

1. When AMTs do not understand the task cards, they ask their supervisor, engineers or quality personnel depending on the task complexity. If necessary, engineers contact manufacturers for clarification of any technical issue related with the aircraft.
2. AMTs have access to a technical aeronautical dictionary to better understand the task cards. Many times the same English technical term is used to reduce potential errors of translation. People learn the meaning of the word from figures and graphs obtained from original manuals.
3. Engineers are responsible for explaining non-routine maintenance procedures to AMTs.
4. Supervisors are responsible for assigning jobs to AMTs according to their technical knowledge and English level. Also for non- routine tasks supervisors often provide on the job training and follow up the AMTs performance.

Chapter 6. EUROPE RESULTS

Only a single country was tested in Europe: Spain. The single site supplied us with 86 participants. With only a single country, many of the analyses have been changed from the earlier regions.

Area	Site #	Number of Employees	Style of Using Task Card in Maintenance
Spain	1	3791	English

Table 6-1. Background information on the MRO in Europe

6.1 Demographics

For each participant we recorded their Gender, Age, Years as an AMT, Years Learning English and Reading Level as given by the Accuracy Levels Test. Table 6-2 shows the mean results for Spain

	Spain
Number Tested	86
Percent Female	2.3%
Age	37.9
Years as AMT	13.0
Yr. Learning English	8.0
Reading Level	5.2

Table 6-2. Demographics of the single country tested in Europe

Compared to Latin America where the same language (Spanish) is used, there were again very few females in the aircraft maintenance population tested. Age was slightly higher, as were Years as AMT. Years Learning English was longer than in Latin American countries, although Reading Level was comparable.

6.2 Incidence Questionnaire

The first analysis was of the overall response to “Have you ever encountered an error of this type?” The incidence of each scenario is shown in Figure 1. Overall, misunderstanding the English documents was the highest frequency of the seven scenarios, followed by Inadequate Verbal Ability. Misunderstanding translations (Scenario 7) was next highest. As in other areas, the three middle scenarios (3, 4 and 5) were much the lowest.

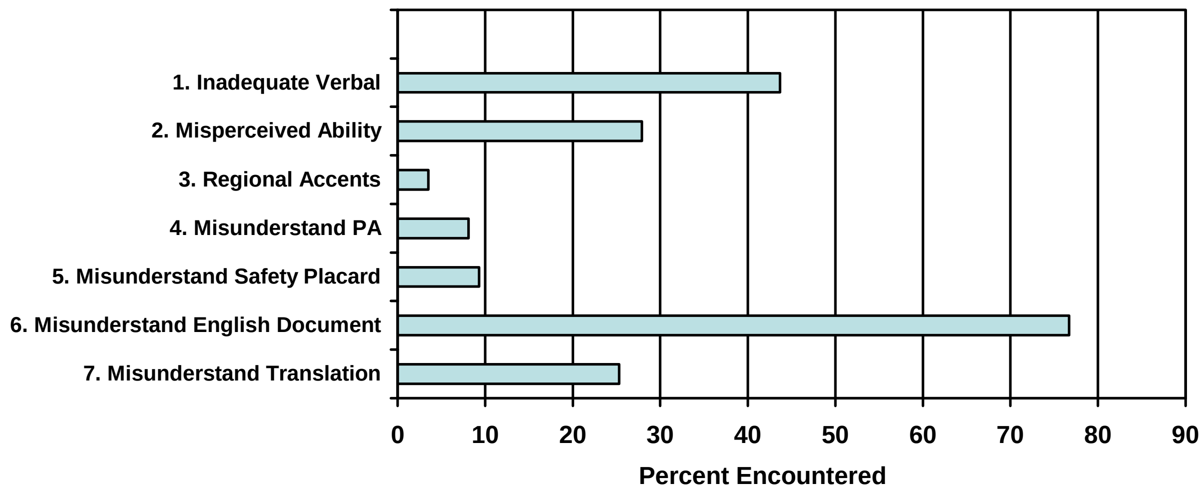


Figure 6-1. Relative frequency with which each of the seven scenarios was encountered

When the answers to the question "When was the most recent time you encountered on errors of this type?" were tabulated, it was possible to estimate the median time since the last occurrence of each scenario. A cumulative plot of probability of occurrence against time since last occurrence for each scenario was used to perform a linear interpolation of the median. The medians are shown for each scenario in Table 6-3 with the mean percentage reported from the previous analysis. Unlike Asia, there was no significant correlation between the two numeric columns of Table 6-3 ($r = 0.170$, $p = 0.715$).

Scenario	Median Weeks Since Previous Occurrence	Mean Percent Reported
1. Inadequate Verbal	12	43.7
2. Misperceived Ability	4.6	27.9
3. Regional Accents	21.6	3.5
4. Misunderstand PA	82	8.1
5. Misunderstand Safety Placard	12.7	9.3
6. Misunderstand English Document	94.5	76.7
7. Misunderstand Translation	207	25.3

Table 6-3. Median weeks since previous occurrence and mean percent reported for each scenario

6.2.1 Error Factors

For the response to factors most associated with these scenarios, GLM ANOVA of the percentage encountering each incident by Factor was performed, with Scenario as the other independent variable. There were no main effects significant at $p < 0.05$. The percentages reporting each factor are shown in Figure 6-2.

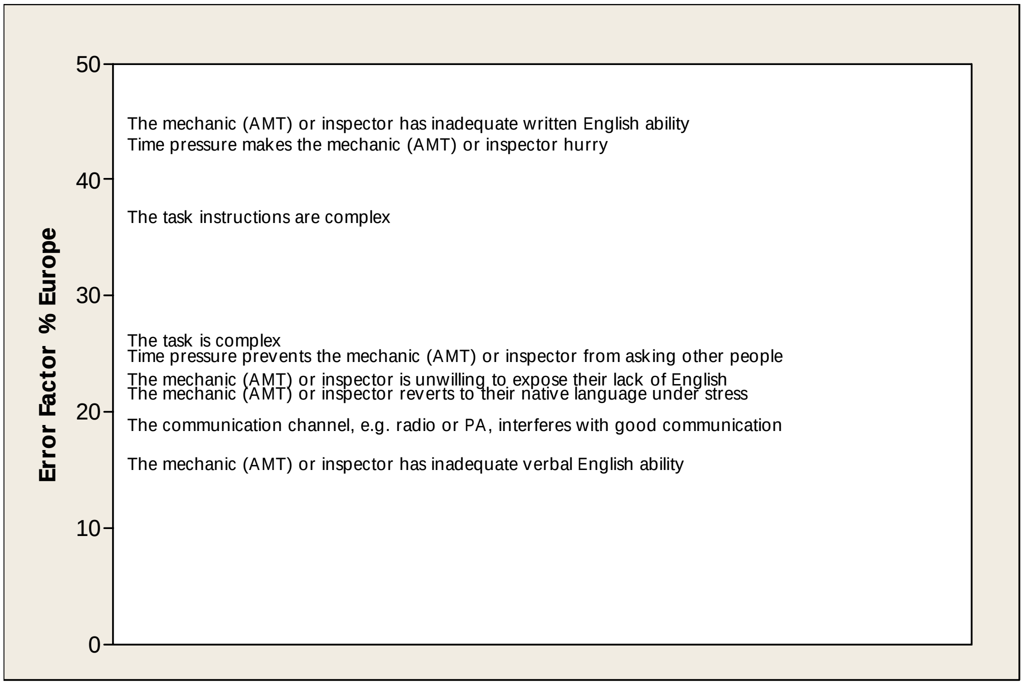


Figure 6-2. Percentage reporting each factor affecting scenario incidence

6.2.2 Prevention Factors

A similar analysis was performed for the ten factors potentially mitigating language errors. The GLM ANOVA gave significance at $p < 0.001$ for Factor ($F(9,54) = 5.9$) and at $p = 0.005$ for Scenario ($F(6,54) = 3.5$). The main effect of Factor is shown as Figure 6-3.

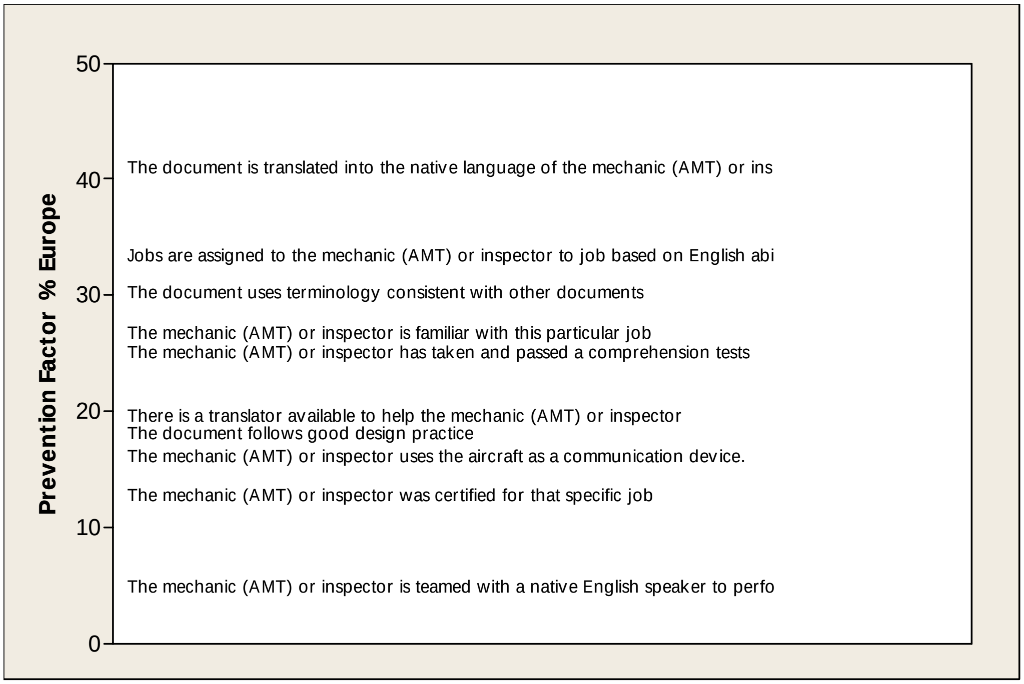


Figure 6-3. Percentage reporting each factor affecting scenario prevention

6.2.3 Discovery Factors

Finally, an analysis of how errors are discovered was performed. The GLM ANOVA gave significance at $p < 0.001$ only for Factors ($F(5, 30) = 22.1$). The main effect of Factor is shown as Figure 6-4.

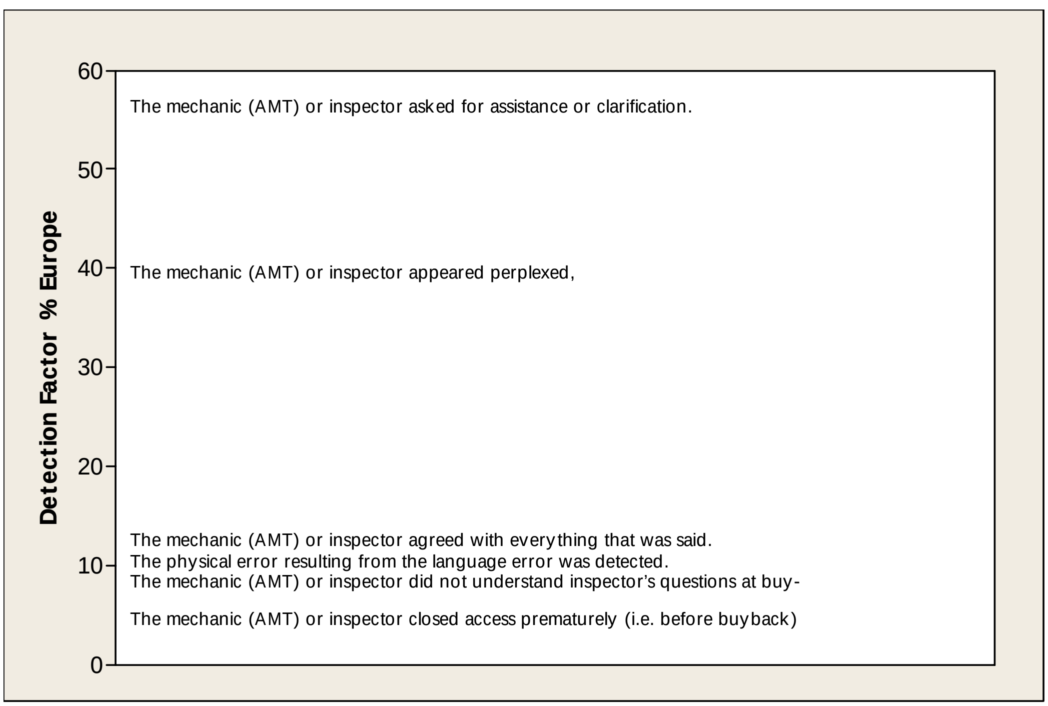


Figure 6-4. Percentage reporting each factor affecting scenario discovery

6.3 Intervention Effectiveness

As in the pre-tests, there was a negative correlation between accuracy (fraction of correct responses) and time (overall time to complete the task) for the comprehension test ($r = -0.233$, $p = 0.030$). This was again a significant speed/accuracy trade off. A third measure was created by dividing Accuracy by Time to give a combined overall Performance score.

Among the demographic variables, there were inter-correlations between Years of Age and Years as an AMT as would be expected, and also a significant correlation of Years Learning English with Reading Level. This latter shows that years spent in English language study does indeed pay off in better reading performance. Another way to express this is that a Factor Analysis (using a Varimax rotation) needed only two factors to explain 82.3% of the variance in these four measures, with the first factor loading Years of Age and Years as AMT and the second loading on Years Learning English and Reading Level. From these analyses of individual characteristics, two relatively orthogonal measures were chosen as potential covariates in the performance analyses: Reading Level and Age.

In contrast to Latin America, there were no correlations of accuracy and time with Age but significant correlations of accuracy and time with Reading Level ($r = 0.224$, $p = 0.037$; $r = -0.336$, $p = 0.001$, respectively).

GLM ANOVAs were performed for each measure (Accuracy, Time, Accuracy/Time) as well as Log_e(Time) because that was found to be more normally distributed than Time. The factors tested were Intervention, Task Card Difficulty and Simplified English, with the two covariates of Reading Level and Age. All main effects and two-way interactions were included except Intervention X Simplified English as the Spanish translations did not differ between Simplified English and non-Simplified English. Higher order interactions were not included due to multiple co-linearity effects.

The results of the ANOVAs are summarized in Table 6-4. With only 86 participants in one country, the ANOVA was simpler, but the results were less significant than earlier analyses. For this reason, some entries in Table 6-4 are classified as significant if they reach a level of 0.10 rather than the usual 0.05 level. The covariate of Reading Level was highly significant in all analyses, but age only reached the 0.10 level.

	Accuracy	**Time**	**Log_e(Time)**	**Accuracy/Time**
Intervention				$F(4,72) = 2.3$ $P = 0.064$
Reading Level (covariate)	$F(1, 72) = 4.7$ $P = 0.033$	$F(1, 72) = 8.9$ $P = 0.004$	$F(1,72) = 9.3$ $P = 0.003$	$F(1, 72) = 10.4$ $P = 0.002$
Age (covariate)		$F(1, 72) = 3.3$ $P = 0.072$	$F(1,72) = 3.5$ $P = 0.066$	$F(1, 72) = 3.8$ $P = 0.054$

Table 6-4. Summary of ANOVA results for intervention performance

To illustrate the predictive power of the covariates, Figure 6-5 shows the four plots of two aspects of performance (Accuracy, Time) against the two covariates (Reading Level, Age). While they show some relationships, the variance is again quite high for all four plots: performance in task card comprehension is more than just good English ability and lower age.

There was only a single effect significant at even the 0.10 level: Intervention for Accuracy/Time. To illustrate this, Figure 6-6 shows both the Accuracy and Time measures. The two translation interventions had higher accuracy with less time taken than the other interventions, although it is emphasized that neither measure alone is significant.

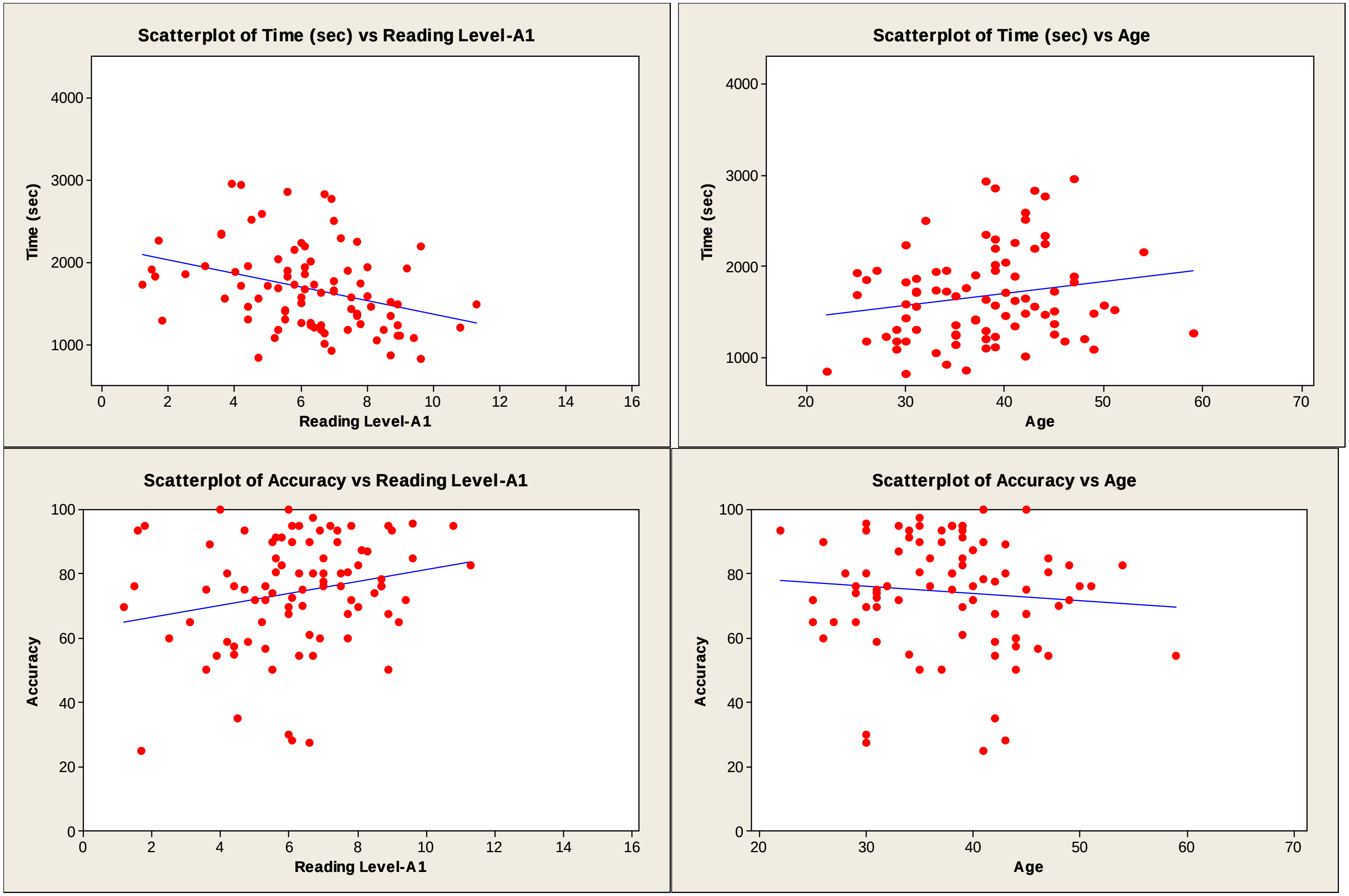


Figure 6-5. Scatter plots of the two aspects of performance (Accuracy, Time) against the two covariates (Reading Level, Age)

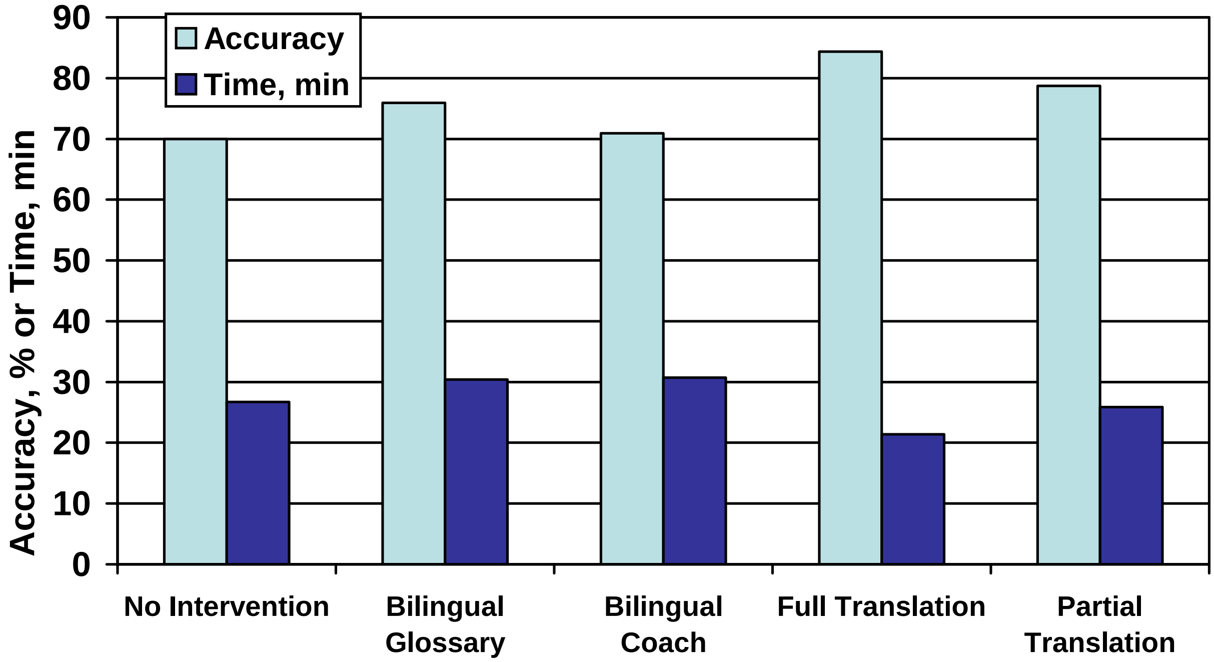


Figure 6-6. Effects of Intervention on Accuracy and Time
(Note: not significant separate effects, see text)

6.4.2 Rating scales

Identical GLM ANOVAs were performed on the fourteen rating scale values, i.e. using Reading Level and Age as covariates and Task Card, Simplified English and Intervention as factors. There were few significant main effects or interactions involving intervention. Figure 6-7 compares the mean scale ratings for all 15 scales, showing that the task cards were generally less highly rated than in other regions. Significant effects (at $p<0.05$) of task card were found for scales 3, 6, 10 and 13, with the Easy Task Card rated more highly than the Difficult one as seen in Figure 6-8. Intervention had a significant effect on Simplicity of English, with the Partial Translation intervention rated much lower (2.0) than all of the others (>4.2). Covariates were only significant for Reading Level on "14. Simplicity of English". For Reading Level ratings of Simplicity of English rating decreased with Reading Level, with a correlation coefficient of $r = -0.268$, $p = 0.015$.

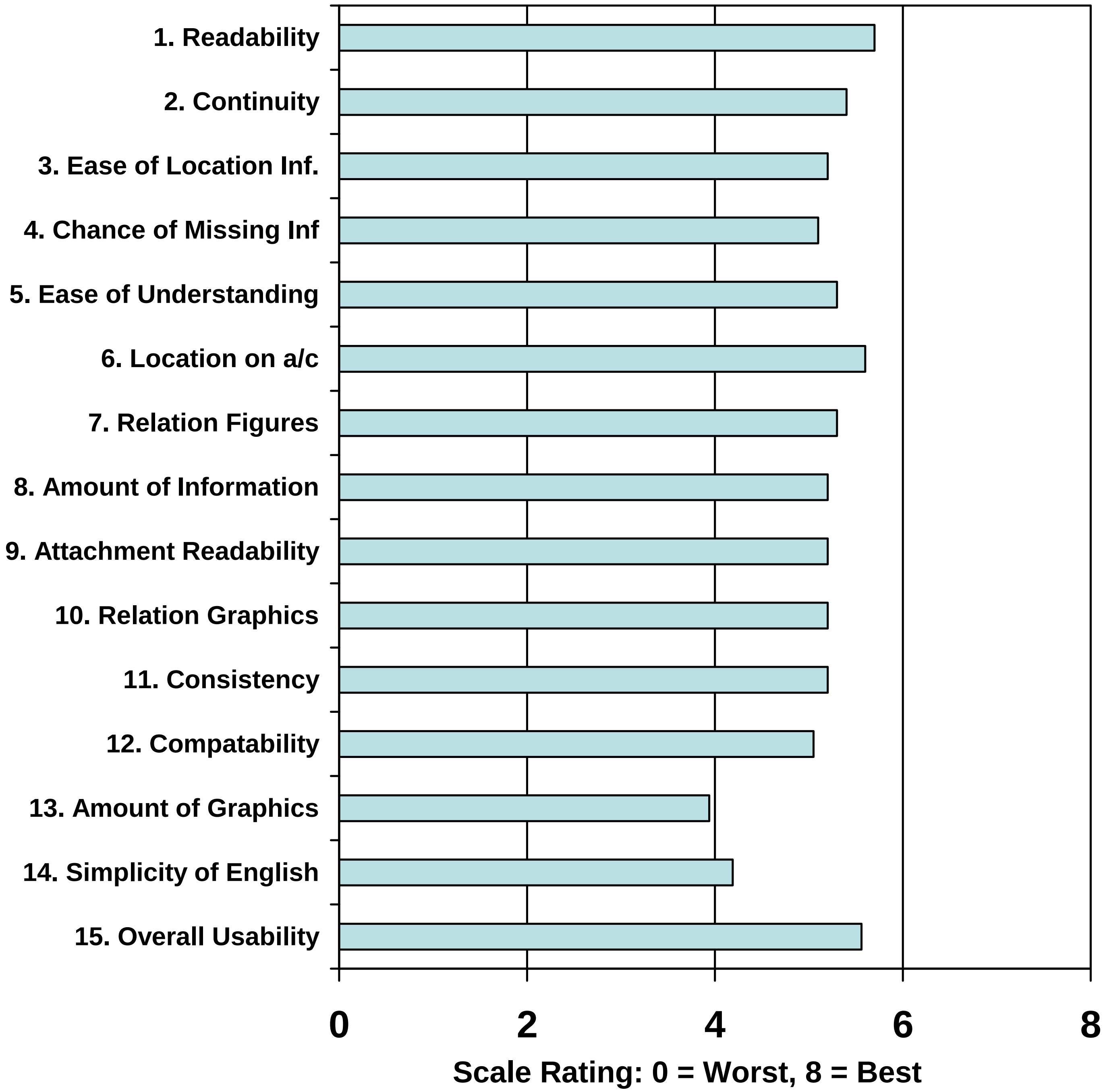


Figure 6-7. Averages of the fifteen rating scales

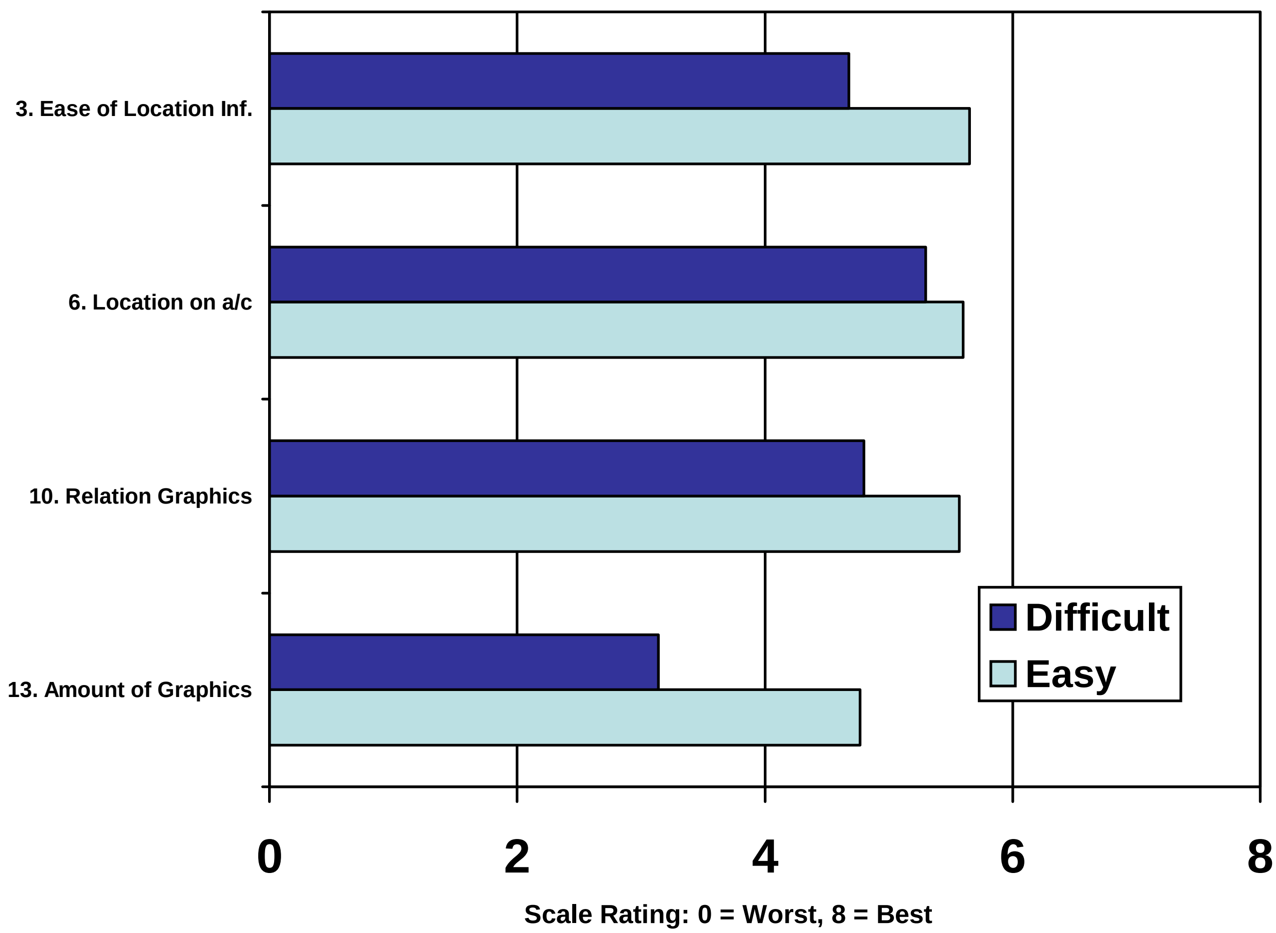


Figure 6-8. Significant effects of task card on four rating scales

6.4 Focus Groups

6.4.1 Current Practice

6.4.1.1 Written communications to AMTs

Ten years ago, all documents were translated to Spanish. By law, pilots needed to have all documentation in Spanish. Many errors were found associated with translated documents, as was the case for Mexico. Currently English is used for task cards, logs and manuals.

6.4.1.2 Writing communications from AMTs

Spanish or English is used depending of the specific document. Some words are not translated and a combination of English and Spanish is used, using English words such as “check valve”, “bleed valve”, “O-ring” and “fan” in Spanish sentences.

6.4.1.3 Verbal Communication

Inside the company the language used is mostly Spanish.

6.4.1.4 Intervention Methods

Better design of documentation: Original English manuals from the OEM manufacturers are used. Engineers need to be able to perform a written translation of any task card. Thus, when hiring new engineers an English test is performed using the maintenance manuals and task cards. Also, a 15 minute verbal interview in English is performed. For non-routine tasks, engineers consult with the different manufacturers to clarify concepts. The English used by French companies (e.g. Airbus) is easier to understand that the English used by American companies. Sometimes, the latter use words that can have different meanings. For some complex task cards, the engineering department translates the whole card except the graphs.
Better education, training and language assistance: Full time engineers are dedicated to do training in technical English and maintenance procedures The Maintenance Training Center (CIM) has five full time instructors and more than eighty part time instructors working for Heavy and Line Maintenance Organization. Airbus and Boeing fleets are both covered at the CIM. Currently more than 140000 training hours are carried out with more than 1100 courses per year. Training courses are designed in accordance with FAA and JAA regulations. Five different types of courses are offered:

1. Familiarization (30 hours); ATA 104 Level I,
2. Mechanics (100-120 hours), ATA 104 Level III.
3. Avionics (130 -150 hours), ATA 104, Level III.
4. Engine Run Up CFM 56 and
5. Troubleshooting, A320/A340.

Three levels of English courses are offered: Basic, Intermediate and Advanced. Attendance on English courses is voluntary and paid by the company. If an AMT wants to be promoted to Supervisor, attendance at English training is taken into account in the decision. All supervisors need to know English well. For clarification, AMTs are expected ask to their lead mechanic, foreman or any engineer.

Chapter 7. USA RESULTS

Two sites of the same company in the USA supplied us with 99 participants. With only a single country, many of the analyses have been changed from the earlier regions.

Area	Site #	Number of Employees	Style of Using Task Card in Maintenance
USA	1	372	English
USA	2	687	English

Table 7-1. Background information on the MROs in USA

7.1 Demographics

For each participant we recorded their Gender, Age, Years as an AMT, Years Learning English and Reading Level as given by the Accuracy Levels Test. Table 7-2 shows the mean results for USA.

	USA
Number Tested	99
Percent Female	1%
Age	43.4
Years as AMT	17.4
Yr. Learning English	n/a
Reading Level	14.3

Table 7-2. Demographics of the USA sample

As expected the Reading Level was very high, a finding similar to other groups of AMTs we have tested (Drury, Wenner and Kritkausky, 1999), about the same as an undergraduate student population. Age, percent female and experience were high and comparable to other groups tested.

7.2 Incidence Questionnaire

In addition to the evaluation of the interventions, we used a questionnaire to determine the relative incidence of the seven scenarios developed earlier. A number of measures of incidence were used, including estimates of the time since last occurrence. The first analysis was of the overall response to "Have you ever encountered an error of this type?" The incidence of each scenario is shown in Figure 7-1. Overall, Inadequate Verbal Ability was the highest frequency of the seven scenarios, with Misunderstanding the English Documents next. As in other regions, the three middle scenarios (3, 4 and 5) were much the lowest, but for the USA Scenario 7 Misunderstanding Translation was also very low in frequency.

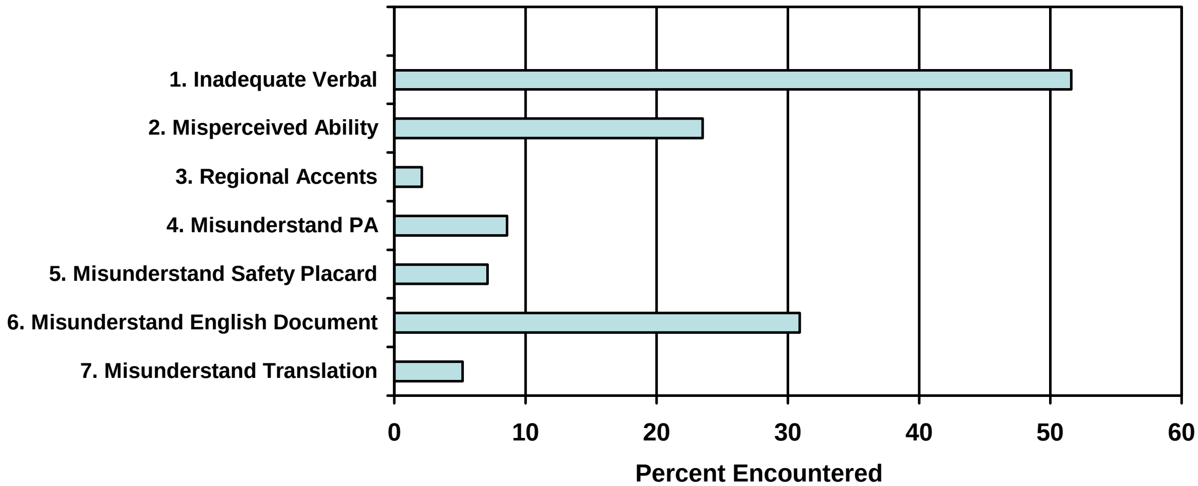


Figure 7-1. Relative frequency with which each of the seven scenarios was encountered

When the answers to the question "When was the most recent time you encountered on errors of this type?" were tabulated, it was possible to estimate the median time since the last occurrence of each scenario. A cumulative plot of probability of occurrence against time since last occurrence for each scenario was used to perform a linear interpolation of the median. The medians are shown for each scenario in Table 7-3 with the mean percentage reported from the previous analysis. Unlike Asia, there was no significant correlation between the two numeric columns of Table 7- 3 ($r = -0.228$, $p = 0.623$).

Scenario	Median Weeks Since Previous Occurrence	Mean Percent Reported
1. Inadequate Verbal	4.8	51.6
2. Misperceived Ability	28	23.5
3. Regional Accents	12.1	2.1
4. Misunderstand PA	1.05	8.6
5. Misunderstand Safety Placard	10.2	7.1
6. Misunderstand English Document	10.5	30.9
7. Misunderstand Translation	28	5.2

Table 7-3. Median weeks since previous occurrence and mean percent reported for each scenario

7.2.1 Error Factors

For the response to factors most associated with these scenarios, GLM ANOVA of the percentage encountering each incident by Factor was performed, with Scenario as the other independent variable. There were no main effects significant at $p < 0.05$. The percentages reporting each factor are shown in Figure 7-2.

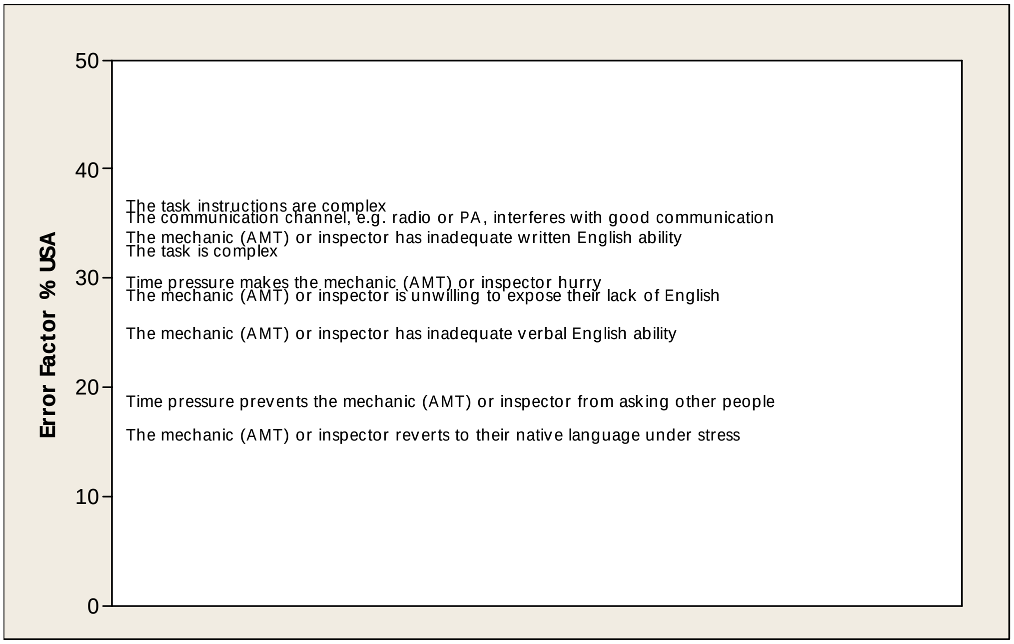


Figure 7-2. Percentage reporting each factor affecting scenario incidence

7.2.2 Prevention Factors

A similar analysis was performed for the ten factors potentially mitigating language errors. The GLM ANOVA gave significance at $p < 0.001$ for Factor ($F(9,54) = 7.6$) and at $p = 0.013$ for Scenario ($F(6,54) = 3.0$). The main effect of Factor is shown as Figure 7-3.

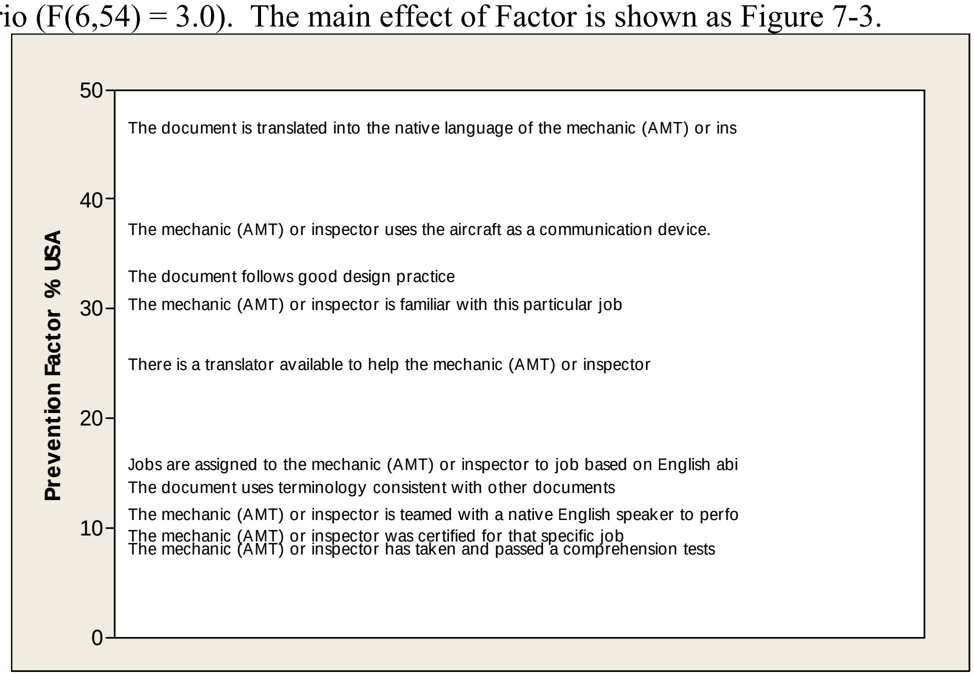


Figure 7-3. Percentage reporting each factor affecting scenario prevention

7.2.3 Discovery Factors

Finally, an analysis of how errors are discovered was performed. The GLM ANOVA gave significance at $p < 0.001$ for Factors ($F(5, 30) = 9.1$) and at $p = 0.023$ for Scenario ($F(6,30) = 2.9$). The main effect of Factor is shown as Figure 7-4.

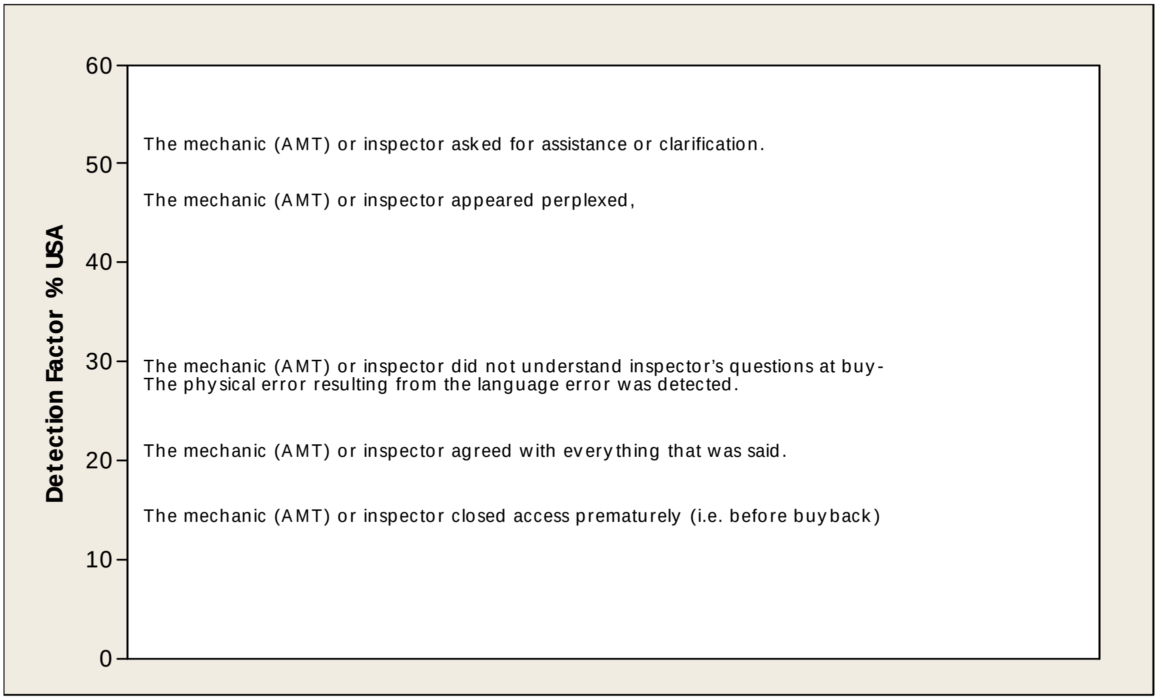


Figure 7-4. Percentage reporting each factor affecting scenario discovery

7.3 Intervention Effectiveness

As in the pre-tests, there was a barely insignificant negative correlation between accuracy (fraction of correct responses) and time (overall time to complete the task) for the comprehension test ($r = -0.196$, $p = 0.051$). This was a possible but not significant speed/accuracy trade-off. A third measure was created by dividing Accuracy by Time to give a combined overall Performance score.

Among the demographic variables, there were inter-correlations between Years of Age and Years as an AMT as would be expected. Note that Years Learning English was not an issue for any of the USA participants tested. Another way to express this is that a Factor Analysis (using a Varimax rotation) needed only two factors to explain 85.1% of the variance in the three measures, with the first factor loading Years of Age and Years as AMT and the second loading only on Reading Level. From these analyses of individual characteristics, two relatively orthogonal measures were chosen as potential covariates in the performance analyses: Reading Level and Age.

There were significant correlations between Age and both Accuracy ($r = -0.212$, $p = 0.040$) and Time ($r = 0.255$, $p = 0.013$). There were also significant correlations between Reading Level and both Accuracy ($r = 0.446$, $p < 0.001$) and Time ($r = -0.309$, $p = 0.002$).

GLM ANOVAs were performed for each measure (Accuracy, Time, Accuracy/Time) as well as Log_e(Time) because that was found to be more normally distributed than Time. The factors tested were Task card Difficulty and Simplified English, with Site as a third factor not expected to show significance, and with the two covariates of Reading Level and Age.

The results of the ANOVAs are summarized in Table 7-4. With only 99 participants in one country, the ANOVA was simpler, but the results were still significant. The covariates of Reading Level and Age were highly significant in all analyses.

	Accuracy	**Time**	**Log_e(Time)**	**Accuracy/Time**
Task Card		$F(1,84) = 14.2$ $P < 0.001$	$F(1, 84) = 14.8$ $P < 0.001$	$F(1,84) = 7.6$ $P = 0.007$
Site X Simplified English	$F(1,84) = 5.7$ $P = 0.020$			$F(1,84) = 4.4$ $P = 0.039$
Reading Level (covariate)	$F(1, 84) = 30.3$ $P < 0.001$	$F(1, 84) = 8.5$ $P = 0.005$	$F(1, 84) = 9.1$ $P = 0.006$	$F(1, 84) = 25.3$ $P < 0.001$
Age (covariate)	$F(1, 84) = 5.8$ $P = 0.018$	$F(1, 84) = 10.3$ $P = 0.002$	$F(1, 84) = 8.0$ $P = 0.006$	$F(1, 84) = 9.5$ $P = 0.003$

Table 7-4. Summary of ANOVA results for intervention performance

To illustrate the predictive power of the covariates, Figure 7-5 shows the four plots of two aspects of performance (Accuracy, Time) against the two covariates (Reading Level, Age). All show significant relationships, but again the between-participant variance is high.

In the USA there were no interventions except Simplified English. This interacted with Site, in the direction of better performance of Simplified English at Site 2 and of Non-Simplified English at Site 1. All effects were less than 9% between extreme conditions so that the result was perhaps a sampling artifact. The task card reached significance for the Time measures and Accuracy/Time. Figure 7-6 shows the expected task card effect of better time performance on the easier task card. The time difference was 21% slower for the Difficult task card.

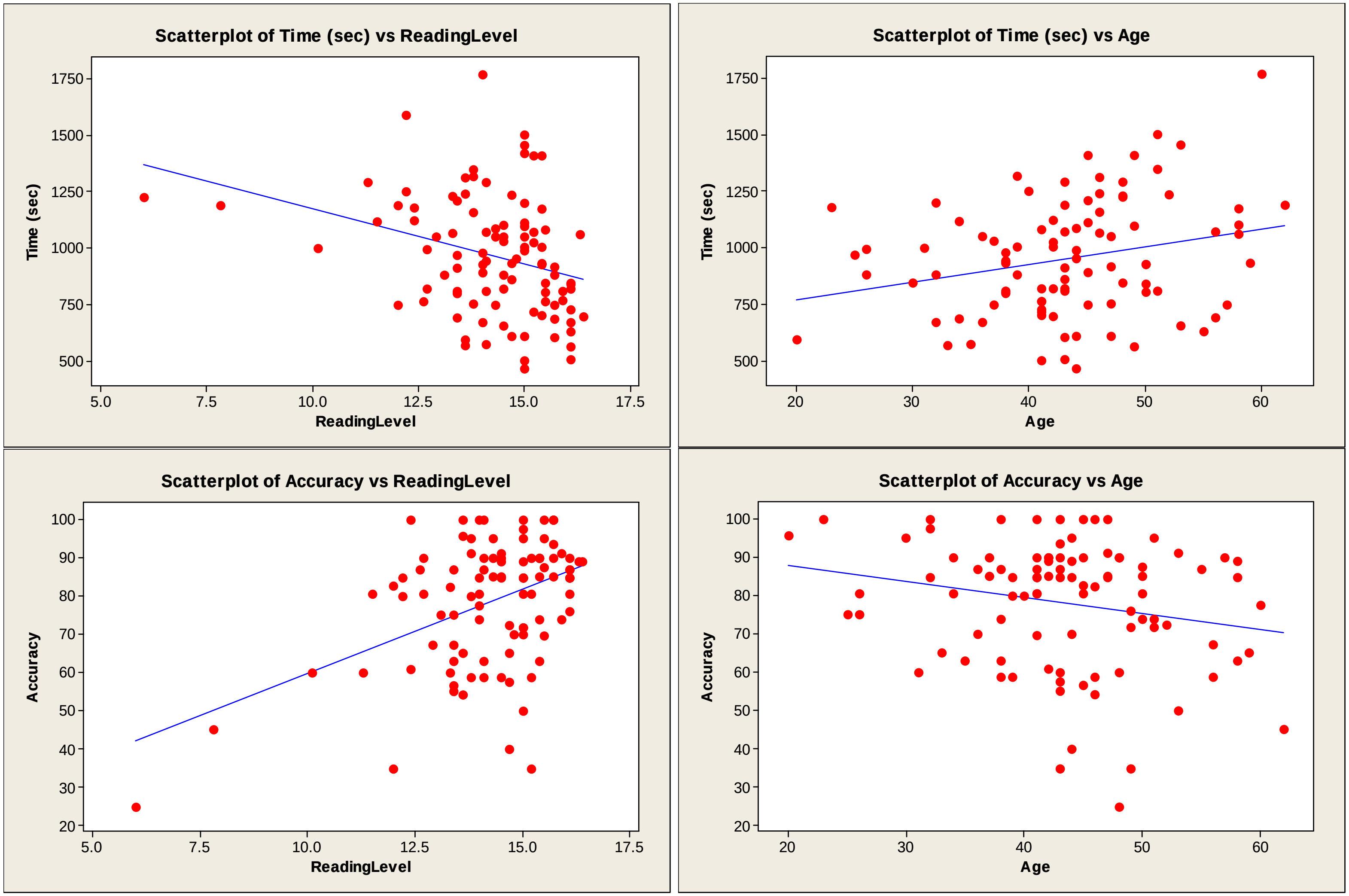


Figure 7-5. Scatter plots of the two aspects of performance (Accuracy, Time) against the two covariates (Reading Level, Age)

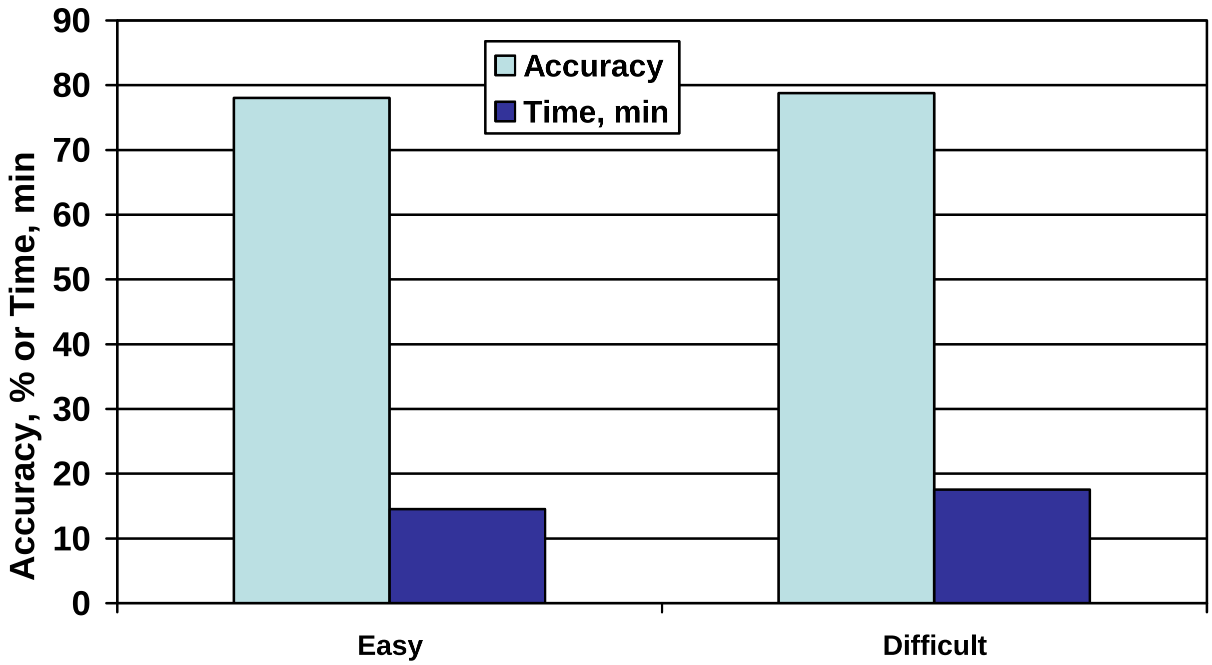


Figure 7-6. Effects of task card on Accuracy and Time

7.3.1 Rating scales

Identical GLM ANOVAs were performed on the fourteen rating scale values, i.e. using Reading Level and Age as covariates and Task Card, Simplified English and Intervention as factors. There were few significant main effects or interactions involving intervention. Figure 7-7 compares the mean scale ratings for all 15 scales, showing that the task cards were slightly more highly rated than in Europe. Significant effects (at $p<0.05$) of task card were found for scale 10 "Ease of Relating to Graphics", with the Easy Task card rated more highly than the Difficult one. Covariates were only significant for Age on "14. Simplicity of English". There were two significant effects of Site X Task Card, for scale 3 "Ease of Locating Information" and scale 4 "Ease of understanding". Neither were large (p values of 0.039 and 0.042, respectively) and again are probably site artifacts.

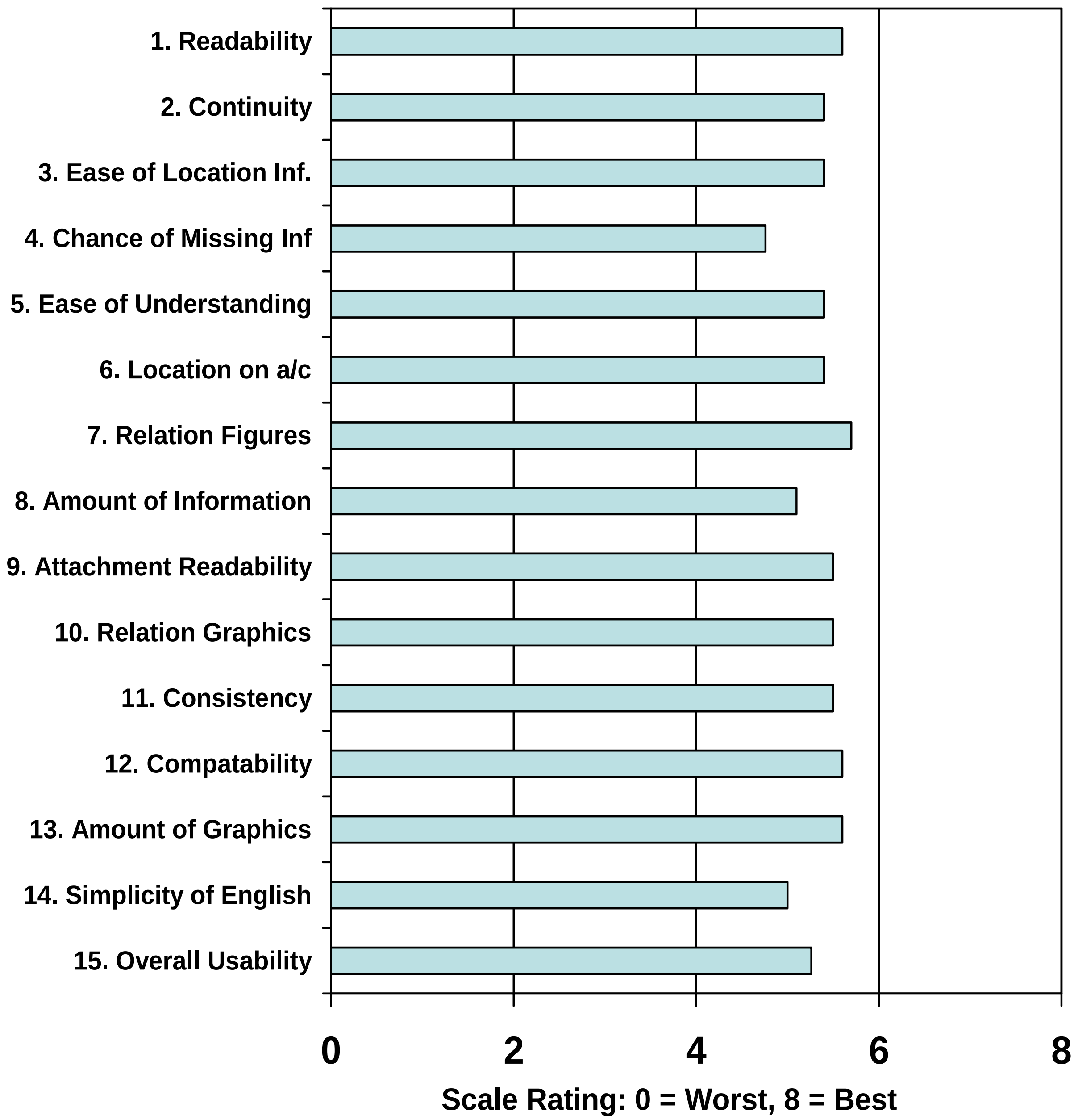


Figure 7-7. Averages of the fifteen rating scales

7.4 Focus Groups

No focus group data was collected for the USA sample as focus groups had already been conducted in the USA as part of the experimental design process (Chapter 1.3).

Chapter 8. INTEGRATION AND DISCUSSION OF RESULTS

This project started by examining the issues in language errors by noting three facts. First is the great increase in contract maintenance among major airlines, now about 50%. The second fact is that the fastest growing segment of the worldwide MRO market is outside the USA. The third fact is that although English is the language of aviation, it is certainly not the native language of most of the world. These facts together produce the logical deduction that language errors may well arise for maintenance of the US civil fleet due to non-native English speakers interacting with maintenance materials in English. Our brief was to assist the FAA in answering a Secretary of Transportation recommendation that:

> *"The FAA should establish a method for determining whether language barriers result in maintenance deficiencies."*

The contribution of this study has been to collect quantitative evidence to test whether this logical deduction is valid, and further to provide quantitative evidence for how any potential problem can be managed.

Fortunately, language error management is not a new problem, only an increasing one. Thus practices have evolved at OEM suppliers, regulatory bodies and maintenance organizations to address language errors. As part of this project, such interventions have been collected and where possible tested quantitatively to guide the aviation community. In the study we have tested 941 aviation maintenance workers on four continents to measure the incidence of language error, to examine the factors leading to such error and its detection, and to measure the effectiveness of chosen interventions.

To ensure that our data collection and analysis procedures were valid, we based them on models of communication, because language errors are a subset of communication errors. The sender and recipient of a message are both important to the final understanding of the message, as is the communications channel. Thus, we expect that the language abilities of those who produce and receive communications will play an important part in language error incidence. The communication channel is also important, whether synchronous as in verbal communication or asynchronous as with most forms of written communication. The channel, such as the PA system or the task card, must be designed as noise free and must transmit the information with minimum ambiguity. Finally, the model includes a feedback loop: the sender needs feedback from the recipient to check that the communication was received correctly. Studies in air traffic control, where bilingual communication was a major issue a decade ago, have shown that recovery patterns from errors often depend on feedback, e.g. requests for clarification or routine use of read-back.

From analysis of an OEM database on language use, we were able to find the prevalence of English and native language usage from 113 airlines around the world. English verbal abilities were highest in North America, followed by Europe and lowest in Asia and the rest of the world. Translation of maintenance manuals was rare while translation of task cards and engineering orders was more common in Asia. Most meetings and training were conducted in the native language. Even among airlines with low reported levels of English ability, translation of documents was uncommon.

Our own observations and focus groups in USA and UK helped us develop a pattern of language error scenarios, and a set of factors that may influence the frequency of these scenarios. While at the sites of these focus groups we pre-tested a methodology for quantifying the effectiveness of language error interventions, for example by providing a bilingual coach, providing an English – native language glossary or using a full or partial translation of a document. The task card comprehension test had been validated in several earlier studies (Drury, Patel and Prabhu, 2000; Drury, Wenner and Kritkauski, 1999) and so was a convenient and useful data collection tool.

From the OEM survey and demographic data on language use, our choice of sites was narrowed to those using a form of Chinese or Spanish. With English, these two languages are the most commonly used on earth. Spanish is spoken in Latin America as well as Europe (Spain) so that we chose as regions Asia, Latin America and Europe (Spain), with a control sample from the USA. While we were measuring scenario frequency and intervention effectiveness, we also collected data on English vocabulary, which gives a direct estimate of reading level on a scale equivalent to US grades in school. Finally, focus groups were held at each site to discuss how that MRO coped with the potential for language error.

Choice of MRO sites within each region was initially based on FAA listings of approved Part 145 sites, which lists the licenses held and the numbers of employees. Sites were contacted individually, usually in their native language, to obtain agreement to visit. Because all the data collection was voluntary, we may not have sampled those organizations that wished to avoid attention. Overall, we were delighted with the level of cooperation at all the sites visited. With over 900 participants in our sample, across four continents, we can examine both the commonalities between regions and regional differences in incidence, intervention effectiveness and reading level.

8.1 Comparisons across Regions

Note that analysis of all of our data can be at the level of the individual MRO site, the country or the region: Asia, Latin America, Europe or USA. Data have been presented in Chapters 4-7 where individual countries and sites have been compared. We now examine regional commonalities and differences.

8.1.1 Reading Levels

Within each region the reading grade levels were typically 4.5 to 5.5 for the samples tested. Higher levels were found where the countries or areas had a history of bilingualism in English: Puerto Rico in Latin America (10.0) and Hong Kong in Asia (6.6). In the USA and England for comparison, Reading Grade levels were very high, about 14, as has been found in earlier studies of AMTs (e.g. Drury, Wenner and Kritkauski, 1999). Overall written English comprehension was at quite a high level throughout: about 5th grade in countries where English is not native or bilingual. The 5-6 grade levels of English reflect an often-stated aim of documentation to be written for a “6^{th} Grade level”, although such a recommendation was never meant to apply specifically to aviation maintenance English.

8.1.2 Scenario Frequency and Factors Affecting

The seven scenarios, developed from our analyses of language error databases and focus groups in the USA and UK, were found to be well-supported in all regions. There were differences in reporting these errors across the countries as shown in the individual region chapters, but consistency across countries was high. A Friedman test of differences between scenario frequencies for the four regions showed a highly significant difference between scenarios ($S(6) = 18.9$, $p = 0.004$), i.e. substantial agreement across regions. This test was used as it is based on ranks of scenarios rather than the absolute magnitude of their frequencies and was thus a fairer test of relative rankings of the seven scenarios. Three scenarios gave high frequencies:

> **Scenario 1:** “The Mechanic (Aircraft Maintenance Technician, AMT) or Inspector was not able to communicate verbally to the level required for adequate performance.”
> **Scenario 2:** “The Mechanic (AMT) or Inspector and the person to whom they were speaking did not realize that the other had limited English ability.”
> **Scenario 6:** “The Mechanic (AMT) or Inspector did not fully understand documentation in English, for example a Work Card or a Manual.”

The most frequently reported scenarios were the ones associated with direct communication surrounding the work itself. All three of these had reported return frequencies between 4 and 10 times per year, and reflected imperfect written communication (work documents) or imperfect verbal communication. The written communication difficulties occurred between the user and English documentation. The examples of scenarios collected from our focus groups confirmed this.

Factors seen as influencing scenario incidence had a large measure of agreement across regions. For Error Likelihood factors a Friedman test similar to the one for scenario incidence was also highly significant ($S(8) = 21.3$, $p = 0.006$), showing high agreement on the relative importance of these factors. There was a consistent group of four highly rated factors:

> The mechanic (AMT) or inspector has inadequate written English ability.
> The mechanic (AMT) or inspector has inadequate verbal English ability.
> The task instructions are complex.
> Time pressure makes the mechanic (AMT) or inspector hurry.

The first two are connected to the individual performing the task, the third is a function of the documentation while the final one is part of the social environment of maintenance.

Prevention factors showed a similar pattern. Again, the Friedman test gave significant factor differences across regions ($S(9) = 22.5$, $p = 0.007$). The five most frequently cited factors that could prevent a language error were:

The mechanic (AMT) or inspector is familiar with this particular job.
The document follows good design practice.
The document is translated into the native language of the mechanic (AMT) or inspector.
The document uses terminology consistent with other documents.
The mechanic (AMT) or inspector uses the aircraft as a communication device, for example to show the area to be inspected.

These again show individual factors, document factors on one procedural factor: using the aircraft as a communication device.

Error discovery factors were also consistent across regions ($S(5) = 18.3$, $p = 0.003$), with just two emerging as highly reported:

The mechanic (AMT) or inspector asked for assistance or clarification.
The mechanic (AMT) or inspector appeared perplexed.

Note that both rely on feedback from the message recipient to the message sender, as our communication model would suggest. Note also that both occur very early in the process: detection of language errors is typically reported well before any maintenance/inspection errors have been committed, or the aircraft is released for service.

The typical picture arising across all of the measures is that language errors of many types are possible, although only a few are frequent, with a language error-prone activity having consistent characteristics:

Complex task instructions
Poorly designed document, in English
Users with low ability in English and low familiarity with the task to be performed
Time pressure to complete the task

When listed in this way, language errors appear to have all of the usual human factors ingredients for error, not just language error. All of these, apart from low ability in English, can be found in standard texts in human factors, such as Wickens and Hollands (2000) as well as those specifically directed at aviation or aviation maintenance (e.g. Garland, Wise and Hopkins, 1999; Reason and Hobbs, 2003) and the *Human Factors Guide for Aviation Maintenance* (Maddox, 1998). The implication is that if the “usual” error-shaping factors are present, then the “usual” interventions should be effective, e.g. training (Taylor, 1993), documentation design (Drury and Sarac, 1997), organization design (Taylor and Felten, 1993; Reason, 1997). We see more evidence for effective interventions as we add the results from the intervention effectiveness experiment and the focus groups.

8.1.3 Intervention Effectiveness

Direct measurement of intervention effectiveness produced significant results, largely consistent across interventions, regions and task cards, i.e. interactions were almost completely absent, making interpretation simpler. First, as expected, Reading Grade level and Age were highly significant covariates across all measures. Younger participants and those with better reading skills performed better, as has been seen in other studies of document comprehension (Chervak and Drury, 2003; Drury, Wenner and Kritkausky, 2000). Such results now extend to a non-native English speaking population. The significant Reading Grade Level correlations show that increasing mastery of English will have a significant impact on comprehension and is a vindication of the English language training programs invested in by many of the MROs we visited.

For the main factors in the intervention effectiveness experiment, the results were again consistent across regions. Intervention effectiveness, measured by comprehension performance, was largely unaffected by anything except some form of task card translation. Surprisingly, Simplified English had no consistent effect, in contrast to our earlier findings that Simplified English was most effective for non-native English speakers (Chervak and Drury, 2003). That finding was for non-native English speakers in the USA, so perhaps SE is less useful when applied in a setting where the native language is other than English. This negative finding appeared for both Chinese and Spanish speakers. Similarly, neither the interventions of a bilingual coach or a glossary produced any significant results, despite their widespread use as interventions at MRO sites. We suspect that at least part of that is due to the fact that almost none of the participants given these interventions used them during the comprehension test. Perhaps people were embarrassed in front of their peers, or did not want to show "weakness" in front of a data collection team from the USA with FAA funding. In hangar floor observations AMTs *did* discuss their work with bilingual supervisors and often produced well-worn English / native language dictionaries. Also note that this experiment was entirely between participants, a safe but relatively low power design in the face of the individual variability shown in the scatter graphs in Chapters 4-7. The fact remains that the only consistent significant intervention was translation.

A direct visual comparison of the effects of translation in different countries and areas is shown in Figures 8-1 and 8-2. No statistical comparison was attempted: our aim is not to measure whether one country is "better" or "worse" than another but to integrate the large mass of data across world regions. These two figures show the accuracy and elapsed time averages across the baseline and translated conditions respectively, averaged across both easy and difficult task cards. Both Simplified English and Non-Simplified English conditions were also averaged for the baseline of Figure 8-1, while Figure 8-2 averaged both full and partial translation where both were included. The USA and Hong Kong did not use translations. To better illustrate these changes, an overlay of these graphs is given as Figure 8-3, with arrows showing changes between baseline condition and translation.

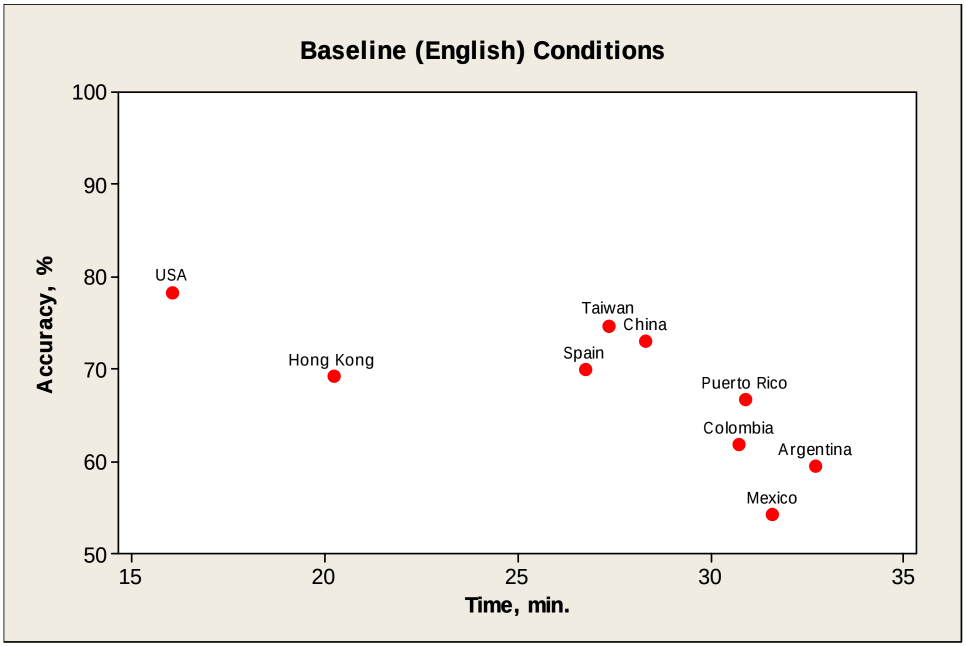


Figure 8-1. Country/area comparison of accuracy and time for baseline condition

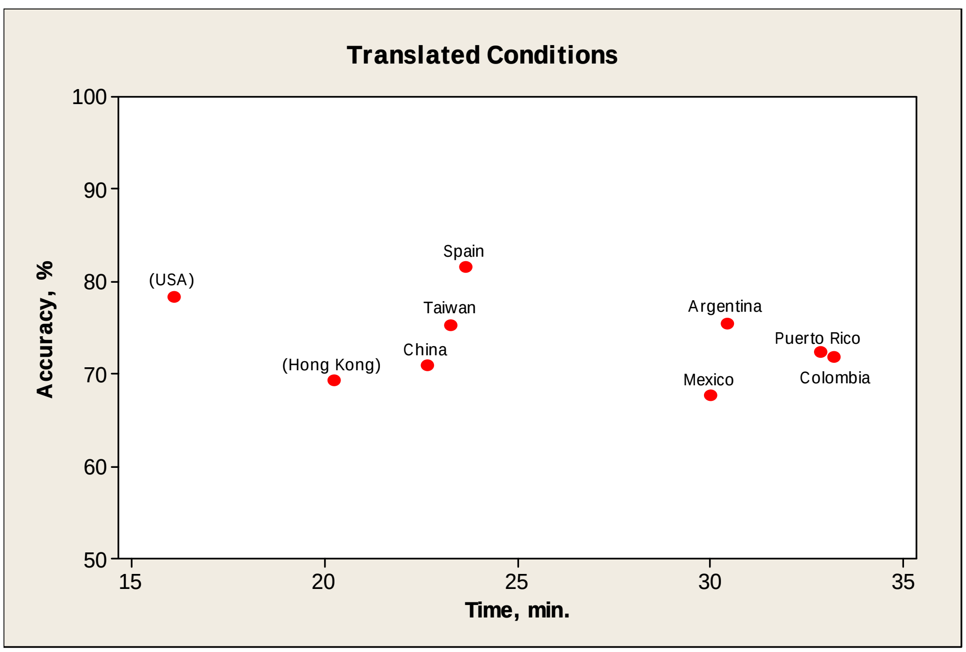


Figure 8-2. Country/area comparison of accuracy and time for translated conditions

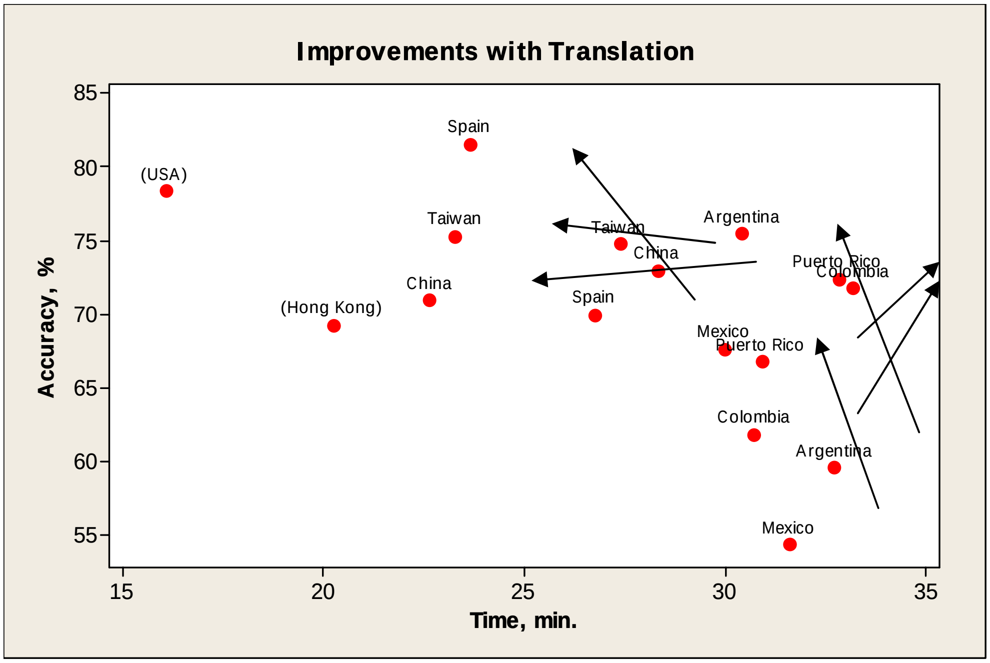


Figure 8-3. Changes in accuracy and time from baseline to translated conditions

Several points emerge from the first of these graphs. First, the USA had consistently the highest accuracy and lowest time: any other result would indeed have been surprising. Second, for the baseline condition, the "best" country or area in each region was the one where bilingualism was the norm: Hong Kong and Puerto Rico. Third, the one European country, Spain, had good performance compared to other Spanish speaking countries, as was expected from the OEM survey results in Chapter 1.

When considering Figures 8-2 and 8-3, a new and interesting picture emerges. First, the accuracy of all countries and areas is now quite comparable, all between about 70% and 80% accurate. [Note that our comprehension test was quite difficult so that 100% would not be expected based on previous results from the USA.] Second, translation has brought up the accuracy performance of all Spanish-speaking countries. At times this was accompanied by an increase in performance speed, while at other times it was not. Third, there was a contrast between Asian countries where translation did not improve accuracy but reduced performance time, and the Spanish-speaking countries where accuracy did improve. In Asia, participants opted for constant (and high) accuracy, letting speed suffer. That is exactly the response the traveling public and regulators would like to see. In Latin America and Spain, accuracy was brought up to a high level by translation, even in Spain where the accuracy was high anyway. From the intervention effectiveness study, the conclusion is that translation works as an error control strategy, bringing accuracy performance to about the same level as in the USA. However, other considerations may be important in choosing translation as an intervention, so the whole issue is taken up later in this chapter.

Overall, the rating scales did not add much of value to the study. Most task cards were rated quite highly and consistently across countries and areas. The main exception was the significant negative impact of translation on task card ratings in Asia: Chinese-speaking participants clearly did not like translated task cards.

8.2 Effectiveness of Interventions

Results from the focus groups covered much ground in wide-ranging discussions, but as our main concern is for effective intervention strategies, we will consider mainly this aspect and integrate results from the other data collection instruments to produce a more comprehensive picture. As an aid to integration, we will use ICAO's SHELL classification of factors affecting human performance: Software, Hardware, Environment, Liveware (individual) and Liveware (social):

8.2.1 Software

This includes both the task itself and the software such as documentation needed to complete the task. Task card complexity was a significant factor in the effectiveness evaluation, with the easy task card usually being completed faster than the more difficult one, although with comparable accuracy. Task complexity was also seen as a major contributing factor in the incidence questionnaire for Asia. Unfortunately, the required tasks in maintenance and inspection are often complex *per se*, but for future aircraft, any help designing inherently less complex tasks would help reduce errors, including language errors.

Documentation is the main issue in language error, both documentation read by the AMT carrying out the work and that generated by the AMT to report progress and completion. Much focus group discussion was on the documentation issue, the complexity and consistency of documentation were factors recognized in the incidence questionnaire, and one documentation intervention (translation) produced a significant improvement in the experimental evaluation, albeit with negative comments from one region. The focus group discussions went beyond the specific issue of task cards, although these are a vital part of any maintenance task. There were issues of wording of source documents, such as maintenance manuals and even the FAR/JARs. After at least 10 years of data on the error-reduction benefits of better documentation (Patel et al, 1994) there is really no excuse for continuing to produce poor source documents. They, like all other job aids, must be designed for the user (AMT) rather than for the convenience of the producer or the dictates of lawyers. There are even design aids validated for error-reduction effectiveness, such as the Documentation Design Aid, DDA (http://hfskyway.faa.gov) to help make the research findings more accessible to busy document writers. Specifically, designers need to use a single word for each concept, provide abbreviation support, use simple sentences and lay out work documentation instructions in an easy-to-follow format. Where the procedure branches, e.g. as a result of an inspection, a flow chart of the procedure is helpful. Within the body of the task card, logical branches should follow a standard format, e.g.

IF (condition)
THEN (procedure step)
ELSE (alternate procedure step)

Many of the above factors are part of Simplified English, which did not prove significant in our experiment. Note, however, that both versions of our documents were well-designed compared with earlier task cards such as those used in the Patel et al (1993) and Patel et al (1994) studies.

All documentation needs to be verified for technical accuracy AND validated by having a representative ultimate user (AMT) perform the task exactly from the instructions. This validation must be by a person outside the documentation design team: just using a document writer with an A&P license is not a validation as such a person knows the task and original engineering documentation too well to act as a naïve user.

While this project has used task cards as the example documentation, this is far from the only form used in maintenance. Across many countries and areas we found the OEM Maintenance Manual and most contracts in English, Engineering Orders and Task cards in both English and the native language, with internal documents such as shift change forms and NRRs mainly in the native language. Training and meetings are also typically conducted in the native language. All of these communication forms must be considered as targets for translation if that might improve performance, despite the current practices.

Translation of task cards was the intervention with the largest potential, both positive and negative. Those sites that used translation fully believed in it, although recognizing its limitations in practice. Those areas not using translation had higher levels of English ability in their user populations (e.g. Hong Kong, Puerto Rico) and saw the errors possible in translation as potential legal traps. Translation did improve comprehension, as expected, but the negative ratings of translated task cards in Asia reflect a dislike of this intervention. In Latin America, translation had been tried several years ago at government insistence, and largely abandoned due to the cost and the need for updates. There was no enthusiasm for returning to this intervention among the focus group participants. One possible modification is partial translation, a well-known strategy in linguistics, where technical terms are left in their original language. In the intervention effectiveness study, we found this equally effective as full translation. It was also observed on the hangar floor as a natural communication medium for AMTs and other technical personnel. Partial translation is less cumbersome than translating technical terms into native languages, where equivalent words may not even exist. It is also perhaps a less time consuming process, although we have no measurements on this aspect.

To improve translation where it is used, the focus group data suggests that the translation be performed by people who know aviation maintenance AND English, not just professional translators or people with degrees in English. Aviation has special uses for words that also have common meanings beyond aviation, and the difference may not be apparent to people without deep aviation knowledge. Unfortunately, AMTs with excellent translation abilities are rare, and perhaps expensive, but any other solution creates the risk of avoidable language error. Whoever is used for translation, consistency is important. The same words must be translated the same way each time, and purely local words should be avoided to ensure that AMTs can move safely between jobs. Standard usage/style manuals should be available and used, as should approved word lists, for example consistent translations from Simplified English. If translation is not used, AMTs need more English language training and practice (see below) but job aids can help. The typical job aids are glossaries and dictionaries, many examples of which were provided to us at the different sites. There is probably a need and market for an aviation maintenance glossary, abbreviation list and dictionary that could be used across sites with the same language. Note however that most languages have variants, e.g. Cantonese *vs*. Mandarin, that need to be accommodated by alternate versions of such a job aid.

Also under Software comes the intervention of protocols. These comprise set words and forms to be used in most communications, modeled on the use of standard protocols in air traffic control and cockpit checklist procedures. Where these are possible, they should be used, for example extending the use of standard English terms as noted above and as implemented in Simplified English.

8.2.2 Hardware

One protocol observed and rated as effective by participants was the use of the aircraft itself as a communications device. This was the primary hardware intervention found where the use of the aircraft (or component) itself aided understanding of the wording of documents. Seeing and touching the aircraft structure has a solid basis in science and represents a good practice in maintenance. This use of the aircraft also requires that the diagrams in the work documents match the structure itself, as seen from the point of view of the AMT, and also including an orientation sketch. Again, these are good practices already well-documented in the literature (Patel et al 1994). Computer-based work documentation may help here as it can provide support for multiple user levels, e.g. good and weak readers of English, by using hypertext format (Drury, Patel and Prabhu, 2000) as well as hypertext links between English and the local language.

8.2.3 Environment

Time pressure was recognized as a factor likely to increase language errors, just as it does other errors. This is well known in the maintenance human factors community (e.g. the Dirty Dozen posters produced by Gordon Dupont) but it still occurs. The issue is not whether it exists, as it probably always will in an industry that tries to maintain schedules despite upsets, but whether the effect of time pressure on errors is recognized by those exerting the pressures. Do managers realize the increased error potential when they demand speed, or when they reward those who "get the job done" while turning a blind eye to cutting corners? After 15 years of aviation maintenance Human Factors Engineering, we still need to ensure that performance-oriented managers (and AMTs who often pressure themselves) consistently choose the "accuracy" side of the Speed/Accuracy Trade-Off (e.g. Drury, 1999).

Liveware (individual): Low English ability, verbal and written, was seen as a causal factor in language error scenarios, a position supported by the significant Reading Grade Level covariate in the intervention study. Most sites recognized this fact and had taken steps to improve English ability of individual AMTs and support personnel. Some MROs had minimum English language entry qualifications, while most had training programs at various levels of ability and tests at each level of promotion. The programs were often elaborate and costly and certainly produced reasonable levels of AMT English usage as measured by the Reading Level. Such programs put a large burden on the individuals involved and their organizations, but appear to be a necessary cost of the historical decision to use English as the only official language of aviation. More use of consistent practice in written and verbal English can help maintain language skills, e.g. NRRs in English or parts of meeting conducted in English. These reinforcements can help prevent a local patois of "Chinese English" or "Spanglish" from taking hold in the organization. The testing of English ability during audits by regulatory bodies is also seen as an effective tool to help ensure an adequate level of English proficiency among AMTs, but it is stressful for AMTs during the audits themselves. When their license and livelihood depend on the audit outcome, stress and even fear is natural. Stress could affect job performance, but checking of English during audits is certainly a necessary part of the quality assurance program for MROs.

Task familiarity was the other individual variable seen as important in reducing language errors. All AMTs start as unfamiliar with each task and develop familiarity with training and repetition. Job assignments should be used to ensure that each AMT becomes familiar with each job in a planned manner, typically starting as an extra hand, then working with an experienced AMT as a coach before performing the task alone. In a busy hangar, this may not always be the easiest short-term assignment arrangement, but it ensures increasing capability over time. As more AMTs are available who are familiar with the task, the scheduling task actually becomes easier over time. In a number of organizations, even job promotion was tied to English ability.

Liveware (social): Human/human interaction is a basic part of Human Factors Engineering, and is intimately related to a cooperative social task such as aircraft maintenance. It also related directly to the language errors we found in this study. Some of the interventions noted above, such as planned task assignment or time pressure, are performed through social means. In the incidence study, the most frequent factor in error discovery was that the AMT asked for assistance or clarification – a social act. The second most frequent discovery factor was that the AMT appeared perplexed, again only important if another person notices. Almost every focus group mentioned that the AMT should ask for language help when needed. From the focus groups also came the need for shift turnovers to be understood across shifts whether in English or native language, the use of engineers (or leads or managers) who better understand English to act as coaches, and discussions between personnel on English interpretation.

The social interventions derived from these factors consist of a number of logical steps. The first is providing some technical English language backup on all shifts, although the Coaching intervention did not give a significant improvement in the intervention study. Unless everybody understands English perfectly (that is not even true in nominally English speaking countries), then having multiple personnel address any ambiguity is preferable to one AMT going ahead and making a best guess and subsequent error. The second is to provide consistency between the documents used for the work and the documents prepared by the AMT or inspector. Shift turnover logs and NRRs should be in English for consistency, unless the organization is very confident in its translations. This intervention will also help reinforce English in a planned manner and give all personnel practice in writing and reading English. Finally, all personnel need to be taught how to recognize when their co-workers are having difficulty with English. An AMT may not want to ask for help (although AMTs not being willing to expose their lack of English was not seen as a causal factor in the incidence study), but co-workers and supervisors should be sensitive to the signs that understanding may be imperfect. As with any social skill, training and practice can help.

Chapter 9. FINDINGS AND RECOMMENDATIONS

This study of language errors in aviation maintenance collected data from 941 participants on four continents: Asia, Latin America, Europe and USA. Seven scenarios for language error were developed, and the current study assessed the incidence of these scenarios. A second study on the same sample was a direct test of the effectiveness of four interventions. Finally, focus groups were used at each site to explore strategies for mitigating language errors. This section brings together the factual findings of the study and recommendations based on these findings. Together, these recommendations form a series of Human Factors Good Practices for reducing the incidence and impact of language errors in aviation maintenance.

Finding 1: The reading grade level of participants at USA MROs was about 14 as found in earlier studies. For MROs on other continents the reading grade level was about 5, with higher levels where there was a history of bilingualism. On all continents, task card comprehension performance improved with reading grade level.

> **Recommendation 1**: Maintenance organizations working on US aircraft should maintain this level of English proficiency among AMTs and other technical staff by using training programs. Training should be aimed at increasing the reading level for English documents.

Finding 2: Of the seven scenarios tested, three where the most common, and had reported frequencies of 4 to 10 times per year:

> **Scenario 1:** "The, AMT or Inspector was not able to communicate verbally to the level required for adequate performance."
> **Scenario 2:** "The AMT or Inspector and the person to whom they were speaking did not realize that the other had limited English ability."
> **Scenario 6:** "The AMT or Inspector did not fully understand documentation in English, for example a Work Card or a Manual."

Most of these language errors were detected early in the process, typically when the AMT asked for help or appeared perplexed.

> **Recommendation 2:** Personnel at MROs and regulatory bodies should be aware of these relatively frequent language error patterns, and take steps from later recommendations to reduce their incidence. Personnel should be aware of the symptoms of these language errors to maintain early detection and mitigation.

Finding 3: The most frequent factors associated with language error likelihood were:

> The mechanic (AMT) or inspector has inadequate written English ability.
> The mechanic (AMT) or inspector has inadequate verbal English ability.
> The task instructions are complex.
> Time pressure makes the mechanic (AMT) or inspector hurry.

And the most frequent factors associated with error mitigation were:

The mechanic (AMT) or inspector is familiar with this particular job.
The document follows good design practice.
The document is translated into the native language of the mechanic (AMT) or inspector.
The document uses terminology consistent with other documents.
The mechanic (AMT) or inspector uses the aircraft as a communication device, for example to show the area to be inspected.

Recommendation 3: These factors should be recognized by MRO personnel and regulatory bodies and a set of good practices in later recommendations should be used to reduce the likelihood of these language errors.

Finding 4: In a comprehension test of task cards, accuracy performance was generally good, and was better in areas that were bilingual. None of the interventions except translation proved effective. Glossaries and bilingual coaches were rarely used by participants even when provided in the comprehension study.

Recommendation 4: MROs should maintain their level of task card comprehension accuracy, and check it periodically using the methodology in the current study. Provision of job aids such as glossaries and bilingual coaches should continue, but AMT should be encouraged to use them.

Finding 5: Translation of documents into a native language is an effective means of improving the comprehension performance. The translation intervention is difficult and costly: it may also be error prone unless done well. However, it did prove effective. In Asia, the improvement was in speed only, but in other regions accuracy also improved. Partial translation proved as effective as full translation where tested.

Recommendation 5: MROs should consider translating documents into the native language, particularly those documents that rarely change, such as task cards and procedures in component shops. Partial translation, where technical terms remain in English, is a good alternative to full translation. The effectiveness of translation may be different in different regions and countries.

Finding 6: Communication design is a key element in reducing language errors. Better communications design includes document design, standard protocols and job aids as well as the English training of Recommendation 1.

Recommendation 6: Better documentation design should be used as standard procedure. This involves using consistent terminology, a single word for each concept and simple sentence structure. Document layout should follow good Human Factors practices, e.g. the Document Design Aid. All documentation should be verified by observing its actual use in the hangar. Job aids such as using the aircraft or component itself in discussing the task, or defining standardized protocols for tasks, should be used.

Finding 7: Other documents and procedures besides task cards need to be considered for language error proneness depending upon their intended user. Non-Routine Repair forms, local contracts and shift turnovers were typically in the native language, while documents for audits, contracts with US companies and maintenance manuals were typically in English. Translation is not the only intervention that needs to be considered. Training and meetings were typically conducted in the native language.

> **Recommendation 7:** While task cards were the specific documents evaluated in the intervention effectiveness study, all other documentation forms need to be examined for their resistance to language errors. Good Human Factors practices in training, communications skills, document design and job aids apply equally to these other documents and situations.

Finding 8: The social environment was found important in language error causation and mitigation. Time pressure on AMTs and inspectors was reported as a major cause of language errors. Regular testing of AMT's English ability, task assignments recognizing the AMT's knowledge of English and an atmosphere of freedom to ask for language assistance were all found important.

> **Recommendation 8:** MROs and regulators should recognize the importance of social interaction factors in language errors, and take steps to ensure that these social factors minimize the incidence of language error.

REFERENCES

Albrecht, T. L., Johnson, G. M., and Walther, J. B. (1993), Understanding Communication Processes in Focus Groups, *Successful Focus Groups, Advancing the State of the Art, edited by D. L. Morgan*, Sage Publications, Inc., Thousand Oaks, CA. p 51-64.

Aircraft Repair and Maintenance Advisory Committee (2001). *FAR Part 145*.
Report to Secretary of Transportation

Cant, M. and Jeynes, L. (1998). What does outsourcing bring you that innovation cannot? How outsourcing is seen – and currently marketed – as a universal panacea. *Total Quality Management*. 9(2/3), 193-202

Carver, R. P. (1987). *Technical Manual for the Accuracy Level Test*. REVTAC Publications, Inc.

Chervak, S. C. and Drury, C. G. (2003). Effects of job instruction on maintenance task performance. *Occupational Ergonomics*, **3(2),** 121-132.

Chervak, S., Drury, C. G., and Ouellette, J. L. (1996). Simplified English for Aircraft Workcards. *Proceedings of the Human Factors and Ergonomics Society 39th Annual Meeting,* 303-307.

Chervak, S., Drury, C. G., and Ouellette, J. L. (1996). Simplified English for Aircraft Workcards. *Proceedings of the Human Factors and Ergonomics Society 39th Annual Meeting,* 303-307.

Davidmann, M. (1998). Using Words to Communicate Effectively, online resource: http://www.solbaram.org/articles/words.html.

Dennis, W. (2002). MRO to Grow in China. *Aviation Week and Space Technology*.

Drury, C. G. (1999). Managing the Speed-Accuracy Trade-Off. In W. Karwowski and W. Marras (eds), *The Occupational Ergonomics Handbook*, NY, CRC Press, 677-692.

Drury, C.G. and Ma, J. (2003). Do Language Barriers Result in Aviation Maintenance Errors? *Human Factors and Ergonomics Society 47th Annual Meeting Proceedings*, Denver, Colorado, October 13-17, 2003

Drury, C.G. and Ma, J. (2003). *Language Errors in Aviation Maintenance: Year 1 Interim Report*, Reports to William J. Hughes Technical Center, the Federal Aviation Administration under research grant #2002-G-025.

Drury, C. G. and Sarac, A. (1997). Documentation design aid development. *Human Factors in Aviation Maintenance - Phase Seven, Progress Report, DOT/FAA/AM-97/xx,* National Technical Information Service, Springfield, VA.

Drury, C. G., Patel, S. and Prabhu (2000). Relative advantage of portable computer-based workcards for aircraft inspection. *International Journal of Industrial Ergonomics*, **26**, 163-176.

Drury, C. G., Shepherd, W. T. and Johnson, W. B. (1997). Error reduction in aviation maintenance. *Proceedings of the 13th Triennial Congress of the International Ergonomics Association'97,* Tampere, Finland, Vol. 3, 31-33.

Drury, C.G., Wenner, C., and Kritkausky, K. (1999). *Development of process to improve work documentation of repair stations.* FAA/Office of Aviation Medicine (AAM-240). National Technical Information Service Springfield, VA.

Drury, C.G., Wenner, C., and Kritkausky, K. (2000). Information Design Issues in Repair Stations. *Proceedings of Tenth International Symposium on Aviation Psychology*. Columbus, OH.

Dupont, G. (1997). The Dirty Dozen Errors in Maintenance. *Proceedings of the 11th Human Factors Issues in Aircraft Maintenance and Inspection Meeting*: Human Error in Aviation Maintenance.

Easterby, R.S. (1967). Ergonomics Checklists: an appraisal. *Ergonomics*, **10(5),** 549-556.

Federal Aviation Administration (2003). "Review of Air Carrier's Use of Aircraft Repair Stations," FAA IG Report AV-2003-047, July 8, 2003.

Federal Aviation Administration (2005). Office of the Inspector General, June 2005.

Fegyveresi, A. (1997). Vocal Cues and Pilot/ATC Communication. *International Symposium on Aviation Psychology*. 81-84.

Flottau, J. (2005). Major MRO Provider Expands. *Aviation Week & Space Technology*, March 14, 2005, Vol. 162(11).

Gardner-Chloros, P. (1991). *Language selection and switching in Strasbourg.* Claredon Press, New York: Oxford University Press, 218.

Garland, D. J., Wise, J. A. and Hopkin, V. D. (Eds.) (1999). *Handbook of Aviation Human Factors,* Mahwah, NJ, L. Erlbaum Associates, 591-605.

Griffith, E. (1999). *A First Look at Communication Theory (4th edition)*. Boston: McGraw-Hill.

Johnson, D.W. (1972). *Reaching Out*. Englewood Cliffs, N.J.: Prentice Hall.

Kanki, B. G., and Smith, G. M. (2001). Training Aviation Communication Skills, In E. Salas, C.A. Bowers and E. Edens. (eds), *Improving Teamwork in Organizations: Applications of Resource Management Training*, Lawrence Erlbaum Associates, Mahwah, New Jersey, 95-127.

Luna, D and Perachio, L (2005). Advertising to Bilingual Consumers: The Impact of Code-Switching on Persuasion. *Journal of Consumer Research*, 31, 4, 760 – 765.

Maddox, M. (1998). The *Human Factors Guide for Aviation Maintenance*.

McAuley, J.G. (1970). *People to people: Essentials of Personal and Public Communication*. Belmont, CA: Wadsworth Pub. Co.

Merritt, A., and Ratwatte, S. (1997). Who Are You Calling A Safety Threat? A Debate On Safety in Mono- Versus Multi-Cultural Cockpits, *International Symposium on Aviation Psychology*, 661-666.

Montes-Alcala, C. (2000). *Two Languages, one pen: Socio-pragmatic functions in written Spanish-English code switching*. Unpublished Ph.D. Dissertation, University of California, Santa Barbara, 265.

Orasanu, J., Davison, J., and Fischer, U. (1997). What Did He Say? Culture and Language Barriers to efficient Communication in Global Aviation, *International Symposium on Aviation Psychology*, 673-678.

Overhaul & Maintenance (2002). MRO in Asia-Pacific Region. Aviation Week's ShowNews Online: www.aviationnow.com. Online resources: *http://www.awgnet.com/shownews/02asia1/mro09.htm*

Patel, S., Drury, C.G., and Prabhu, P. (1993). Design and usability evaluation of work control documentation. *Proceedings of the Human Factors and Ergonomics Society 37th Annual Meeting, 11*56-1160, Seattle, WA.

Patel, S., Drury, C. G. and Lofgren, J. (1994). Design of Workcards for Aircraft Inspection. *Applied Ergonomics*, **25(5)**, 283-293.

Phillips, E. H. (2004). MRO Mired in Red Ink. *Aviation Week & Space Technology*, January 19, 2004, Vol. 160(3).

Phillips, E. H. (2005). Maintenance Repair and Overhaul. *Aviation Week & Space Technology*, May 2, 2005, Vol. 162(18).

Reason, J. (1990). *Human Error*. Cambridge University Press.

Reason, J. (1997). Approaches to controlling maintenance error. *Proceedings of the FAA/AAM 11th Meeting on Human Factors Issues in Aviation Maintenance and Inspection*, San Diego, CA.

Reason, J. and Hobbs, A. (2003). *Managing Maintenance Error, A Practical Guide*. Burlington, VA, Ashgate Publishing.

Rifkind, L. (1996). *Communication, Human Factors Guide for Aviation Maintenance*. Washington, DC, Federal Aviation Administration/Office of Aviation Medicine, Online resource: http://hfskyway.faa.gov

Sparaco, P. (2002). French foresee broader, multinational MRO network. *Aviation Week and Space Technology,* 72-73.

Seidenman, P. and Spanovich, J. (2004). European and Asian MRO Rebounding and Expanding. *Aviation Week & Space Technology*, September 13, 2004, Vol. 161(10).

Taylor, J. C. (1993). *The effects of crew resource management (CRM) training in maintenance: an early demonstration of training effects on attitudes and* performance. DOT/FAA/AM-93/5. Human Factors in Aviation Maintenance - Phase Two Progress Report, Alexandria, National Technical Information Service.

Taylor, J. C. and Felten, D. F. (1993). *Performance by Design*. NJ, Prentice Hall.

Taylor, F. W. and Patankar, J. (2000). The Role of Communication in the Reduction of Human Error. *Proceedings of the 14th International Symposium on Human Factors in Aircraft Maintenance and Inspection*, Vancouver, B.C.

Wall, R. (2005). Overhauling Maintenance. *Aviation Week & Space Technology*,April 18, 2005, Vol. 162(16).

Wenner, C. A., and Drury, C. D. (2000). Analyzing Human Error in Aircraft Ground Damage Incidents. *International Journal of Industrial Ergonomics*, 26, 177-199

Wickens, C. D. and J. G. Hollands (2000). *Engineering Psychology and Human Performance*. NJ, Prentice Hall.

Wideman, R.M. (2002). *Communication Models*. Online resource: http://www.maxwideman.com/issacons4/iac1432/